AD_____

Award Number: DAMD17-00-1-0306

TITLE: Reconstruction of Mammary Gland Structure Using Three-Dimensional Computer-Based Microscopy

PRINCIPAL INVESTIGATOR: Carlos Ortiz de Solorzano, Ph.D.

CONTRACTING ORGANIZATION: University of California at Berkeley
Berkeley, California 94720

REPORT DATE: August 2004

TYPE OF REPORT: Final

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 074-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	August 2004	Final (1 Aug 00 - 31 Jul 04)

4. TITLE AND SUBTITLE
Reconstruction of Mammary Gland Structure Using Three-Dimensional Computer-Based Microscopy

5. FUNDING NUMBERS
DAMD17-00-1-0306

6. AUTHOR(S)
Carlos Ortiz de Solorzano, Ph.D.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
University of California at Berkeley
Berkeley, California 94720

E-Mail: Mc131367@bcm.tmc.edu

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES
Original contains color plates. All DTIC reproductions will be in black and white.

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 Words)

During the administrative funding period of this grant we have developed a system that permits three-dimensional reconstruction of entire the entire murine ductal epithelium from physical tissue sections. The system reduces the interaction required for low-resolution imaging of H&E stained sections, registration of images of consecutive sections and annotation of tissue structures -i.e. morphologically normal ducts and intraductal tumors- in the images. In addition, we have developed fully automatic tools for image registration and annotation that are now being integrated in our system and used in the reconstruction of the latest tissue specimens. Complementing morphological H&E based reconstruction, our system can be used for morphologically driven acquisition of high-resolution images from immunostained intermediate sections, both using fluorescence and brightfield microscopy. To show the use of our system, we have used it to compare the whole-gland distribution of hormone receptors (ER and PR) in the development of the mouse mammary gland. For that, we have imaged and quantified the whole-gland cell-by-cell ER and PR status of the mammary gland as it evolves through puberty to maturity and into menopause.

14. SUBJECT TERMS
Three-dimensional microscopy, computer-aided tissue reconstruction

15. NUMBER OF PAGES
74

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

Table of Contents

INTRODUCTION (as in the original proposal)

Understanding normal biology and its pathological deviations requires comprehending the form in which the different tissue components work together in their native tissue context. In particular, it requires studying cell-to-cell and cell-to-environment interactions, which tend to be highly variable between different tissue parts. This information can only be acquired by studying tissue in a state as similar to its native context as possible.

Contextual information, which is important for understanding normal tissue behavior, becomes essential when studying Breast Cancer, since heterogeneity and three-dimensionality are at the very base of cancer initiation and clonal progression. However, many analytical procedures in biology start by taking the element under study (RNA, DNA, etc) outside its native context. By doing so, an accurate measure of that element is achieved, but priceless information on how that element is related to its environment is lost. Other methods based on immunohistochemical staining of thin tissue sections for visual observation at the conventional microscope preserve only partial contextual information, since they provide a flat -two dimensional- view of a three-dimensional object.

Our goal is to develop a three-dimensional computer-based quantitative microscopy platform to be applied to simultaneous morphological and molecular studies of both normal and neoplastic mammary tissue. Using our system, we will be able to reconstruct relevant microscopic tissue structures (e.g. mammary ducts, lobules, lymph nodes, tumor masses, etc.) from consecutive thin tissue sections. Then, using the 3D virtual reconstruction, it will be possible to perform quantitative morphological measurements combined with directed –structure defined- high-resolution analysis of molecular events.

To show the use of our system, we will study the combined role of Estrogen and Progesterone Receptors (ER & PR) in mammary gland development. We will quantify cell-to-cell the levels of ER and PR in different parts of the mammary gland, at different time points of the development of the gland.

BODY

Our accomplishments for the entire funding period (08/01/00-07/31/04) will be described following the Tasks enumerated in the approved proposal.

As mentioned in last year's "final" report, the technology developments required for this grant were done under the join budget of this and another grant, "*Mechanisms of Intraductal Tumor Spread*" (DAMD17-00-1-0227). That is why the technology accomplishments reported here (mainly Tasks 1 and 2) are similar to those described in the other grant's report.

Note: this report corresponding to the one-year no cost extension requested last year has been written by completing the "final" report of the grant sent in August, 2003. We have kept the statements relative to the accomplishments done during the first three years, adding those obtained during the one-year no-cost extension.

Task 1. (Months 1-12) Modify an existing microscopic imaging system for acquiring low magnification (1 pixel= 5 μm) images of entire tissue sections and for tracing in 3D the ducts in the tissue specimen from a series of images of adjacent sections.

1. Complete the existing JAVA based software for interactive marking and 3D virtual rendering of ducts so that it allows any branching pattern. (Months 1-6)
2. Develop a customized VRML viewer to allow visualization of the branching pattern and hyperlink it to the original images (Month 6-12).

Task 2. (Months 6-30) Interface the existing acquisition and registration software with the JAVA application to allow revisiting of acquired slides for inspection and high-resolution acquisition of areas of interest. (Months 6-12)

1. Integrate the JAVA application with the conventional fluorescence microscope (Months 6-12)
2. Integrate the JAVA application with the confocal microscope (Month 12-18)
3. Integrate the high-resolution images in the VRML 3D representation of the mammary gland (Months 12-20)
4. Integrate the image analysis methods for nuclear segmentation with the JAVA application (Months 18-24)
5. Integrate the results of the segmentation with the VRML representation of the mammary gland. (Month 24-30)

At the end of the entire funding period, we can present R3D2, a robust JAVA based software that can be used to semi-automatically image and reconstruct tissue structures from serially sectioned thin (5 μm) tissue sections. The reader can find a complete description of the system can be found in Appendix 1, a reprint of a paper published in the journal Microscopy Research and technique (Bibliography J1). What follows highlights the main features of this system.

This system controls a fully automated scanning microscope and can automatically acquire entire sections stained for either fluorescence (e.g. DAPI) or bright field (e.g. H&E) microscopy. The system acquires and tiles together multiple single field-of-view images that cover the entire extent of the section, adjusting the focus plane of the microscope, when required to maintain optimum contrast of the acquired images (see Appendix 1 Fig. 4). All the related sections that make up a tissue block are stored following a predefined directory structure and can be loaded and browsed easily. To allow real-time loading and browsing of these extremely large images (some of them can be up to 120Mb in size) we have used memory mapping techniques that permit accessing the large image files without loading the entire image pointer in the computer memory. Visualization of the entire section is done by displaying reduced version of the original imaged, subsampled to fit the size of the visualization window. Then, zooming in areas of the section is done by retrieving the image data directly from the image file, based on image coordinates selected in the subsampled version of the image of the entire section.

Images can be acquired in grayscale using the 12 bit CCD camera attached to the microscope, or in color by sequentially acquiring three grayscale images using the proper excitation/emission filters (fluorescence) or an RGB tunable liquid crystal filter (brightfield). Visualization of 12 bit images in a 8 bit per color depth display can be done by linear compression or truncation of the image data. In addition, images can be linearly stretched to enhance low contrast images.

R3D2 provides interactive tools for registering images of consecutive sections, which is necessary to correctly render in three dimensions tissue structures traversing multiple sections. This is critical, due to the significant misalignment between sections, consequence of the manual sectioning process. In addition to linear effects (shifting and rotation), there are frequent non-linear effects caused by tissue folding, shrinking, stretching or tearing which make registration extremely difficult. Within the scope of this

grant, we have also solved and automated the linear registration problem and we have proposed a solution to correct problems due to non-linear effects (see below).

R3D2 is equipped with annotation tools (active pencil, selection, grouping, splitting, etc) that can be used to manually delineate and connect (within and between sections) tissue structures (in our application normal ducts and lobulo-alveolar structures, DCIS tumors, areas of invasive carcinoma, etc. see Appendix 1 Figs. 5,6), that can be rendered in 3D using a modified Delaunay triangulation and the surface rendering OpenGL-based toolkit embedded in Java3D. As shown, Java ensures seamless continuity between image acquisition, annotation and reconstruction (rendering). The 3D rendering of the tissue is interactive, in that the entire tissue "scene" can be seen from different view points, at the distance and angle of the user's choice. Also, specific tissue "volumes" can be selected to retrieve information (see Appendix 1 Figs. 7,8).

Finally, R3D2 allows revisiting the original tissue slides from both the images of the entire sections or from the 3D reconstruction of the tissue. Revisiting can be uses to visually inspect the slide under the same of different optical conditions (i.e. lense magnification, light-filtering, etc) although is normally used in to acquire multiple-color areas of interest at high magnification.

A detailed description of the system, extending what has been summarized above, has been published in the journal "Microscopy Research and Technique" (Bibliography J1 –Appendix 1-), and before publishing, presented in International conferences (Bibliography P1, A1, A2). Some news agencies covered and distributed new releases that were published in several newspapers and online news services (Bibliography O1, O2, O3, O4).

Following the described developments, and in the process of applying R3D2 to the reconstruction entire, fully sectioned mammary glands, it became clear that new technical developments were necessary to increase the throughput of the system, since the time required to reconstruct one case using the existing mostly manual tools was in the order of two months. The two main bottlenecks in the process described above are the manual registration and annotation of the sections. Therefore we developed automated tools for both tasks, which have been integrated in the system and are currently being used, providing substantial time savings.

We developed a multiscale, multiresolution registration algorithm based on gradient correlation between consecutive image sections. See Appendix 3 (Bibliography

P3) for a full description of the algorithm. What follows is a brief description of it. The algorithm calculates the optimum *rigid body* transformation (rotation plus translation) between each two consecutive images. To reduce the computational cost, we start using heavily subsampled images, and refining the registration using decreasingly subsampled images. At each iteration level, the registration is calculated as follows: first, the gradient of both images (reference and registering) is calculated and thresholded using an adaptive threshold. Then the distance transform of the images is obtained, that contains the distance of each point to the nearest gradient area (i.e. boundary). Finally the distance transform corresponding to the image being registered is scanned over that of the reference image, for different rotations and translations, and the optimum registration is defined as the absolute minimum of the product of both images, which ideally corresponds to 0, i.e., to perfect overlap between both sections. This method has been presented at an international conference (Bibliography A3) and accepted for platform presentation at another (Bibliography P3).

To address the second bottleneck, which is the annotation of histological structures we used partial differential equation (PDE) morphologically driven flows (i.e. Level Set methods). See Appendix 2 (Bibliography J3) for a full description of the method, or read the summary in the following paragraphs.

A description of the Level Set (LS) methodology is out of the scope of this report. Therefore, a very succinct user-focused is provided next. In a nutshell, the LS approach considers the image as a force or energy field determined by one or a combination of selected image features (e.g. intensity, gradient, object curvature, distance...). Then the segmentation of objects is done by letting some initial seeds manually placed on the original image evolve under the driving force of a velocity function that depends on the energy field. This way, assumed that the right energy field is selected, the curves (surfaces in 3D) that define boundaries of the seeds will converge in or near the boundaries of the objects that one wants to extract.

The initial curve is represented here as the zero level set of a higher dimensional function, and the motion of the curve is embedded within the motion of that higher dimensional function. The speed of that motion is defined based on the characteristics of the image to be segmented. In our case the speed is adjusted so that when the interface is on top of areas with a low gradient it expands quickly, whereas when the gradient is large (indicating the location of an edge in the image) the curve is slowed down. In addition, a surface tension term is included in the speed function to slightly retard or

accelerate the contour depending on its curvature, thereby preserving the smoothness of the advancing front. This approach offers several advantages. First, the zero level set of the higher dimensional function is allowed to change topology and form sharp corners. Second, geometric quantities such as normal and curvature are easy to extract from the hypersurface. Finally, everything expands directly to three dimensions if we embed the advancing three-dimensional surface as the zero level set of a four-dimensional function.

However, changing an n-dimensional problem into one in n+1 dimensions increases the computational cost associated with the method. The narrow-band approach accelerates the level set flow by updating the position of the curve only in a narrow vicinity of its current location. But in our experience, the narrow band technique does not reduce the computational cost to a reasonable limit, due to the large size of the images of the sections. Thus, we propose to use the fast marching method, a numerical technique to solve the equation that drives the movement of the curve by combining an efficient –constrained- solution to the equation of the movement of the front, narrow-band level set methods and a min-heap data structure. This method is only used for monotonically advancing fronts (speed always positive or negative), providing a result very fast, albeit not as accurate as the one obtained by using the level set algorithm. This result is then used as the initial condition for the slower but more accurate level set segmentation.

We have used this approach to segment histological structures on H&E and fluorescent (counterstained) sections, by placing seed points either in the background or in the lumen (See Appendix 3, Figs. 5,6,7,10 for some results) The results have been published this year in the peer reviewed conference proceedings of the SPIE Biomedical Optics Conference in San Jose, CA, (Bibliography P2) and a full paper describing the method and results has been published in a special issue on Breast Cancer of the Journal of Biomedical Optics. (Bibliography J3 –Appendix 2-).

New tools have been also developed during this last year, that have to do with the annotation of high-resolution images of areas of immunostained sections and with the spatial analysis of cellular events in tissue sections. On one hand, new interactive tools have been incorporated to R3D2 to allow annotation of molecular status in images of immunostained cells. The tools allow annotating each individual nucleus of an image, assigning them a level of protein expression (Negative, Low, High) relative to one or more than one (when using double or triple immunostaining) protein. There are other tools to delete the annotations, change them, etc, along with tools to mask and classify

areas of the images. We have used these tools (see below) to classify epithelial areas of the mammary gland of mice, as being growing end buds, characteristic of pubertal gland development, small or large ducts or alveolar structures. After classifying the areas, a comparative analysis of the expression of proteins can be done, to compare the expression in different, morphologically distinct areas of the gland. (see below). In addition, tools have been added to study the spatial statistics of cellular distribution (and pattern of protein expression) in the high resolution areas. These areas will be used to determine patterns of expression (grouping, rejection, etc) between cells expressing specific proteins, or between cells expressing or not several proteins (when using multiple immunostaining). See Appendix 4 (Bibliography P4) for a detailed description of the spatial analysis. Our method has been accepted for platform presentation at an international conference (Bibliography P4), and will be submitted shortly to a peer reviewed journal.

Finally, we have done work to improve the results of the registration. As it has already been mentioned, the manual tissue sectioning can produce non-linear deformations in the tissue, such as folding, stretching, tearing. Occasionally, due to the tissue conditions or improper maintenance of the microtome, some sections are damaged beyond recovery and are disposed, introducing gaps in the sequence of sections that make up the case. All these effects may cause large misalignment between areas of the section that cannot be corrected for solely by applying the global affine transformation previously described. Many non-rigid registration methods have been already proposed in the literature. Specific solutions based on complex transformations, such as elastic registration or piecewise registration, are too expensive in computation time given the dimensions of the images of our histological sections. Therefore we opted for a local registration solution that can produce accurate results in a reasonable time.

Our local registration algorithm divides the reference image in sub-images and calculates a correction vector for each sub-image. The correction vector is calculated using the cross-correlation between the sub-image in the reference image and the correspondent sub-image in the target image. This target sub-image is defined from the rigid registration parameters, if a previous rigid-body registration has been previously applied to the image. After calculating the correlation, every pixel in each sub-image is applied the appropriate translation in the x- and y- axis defined by the correction vector.

The correlation between sub-images can be efficiently calculated in the frequency domain. The result is two new images: modulus and phase. The modulus of the correlation has a peak located a distance that corresponds to the translation of one of the images that would cause the best alignment or similarity with the other image. Therefore, the shift vector is obtained by calculating the difference between the coordinates of the center of the image and the coordinates of the brightest peak in the modulus image . A graphical description of this process is shown the following Figure.

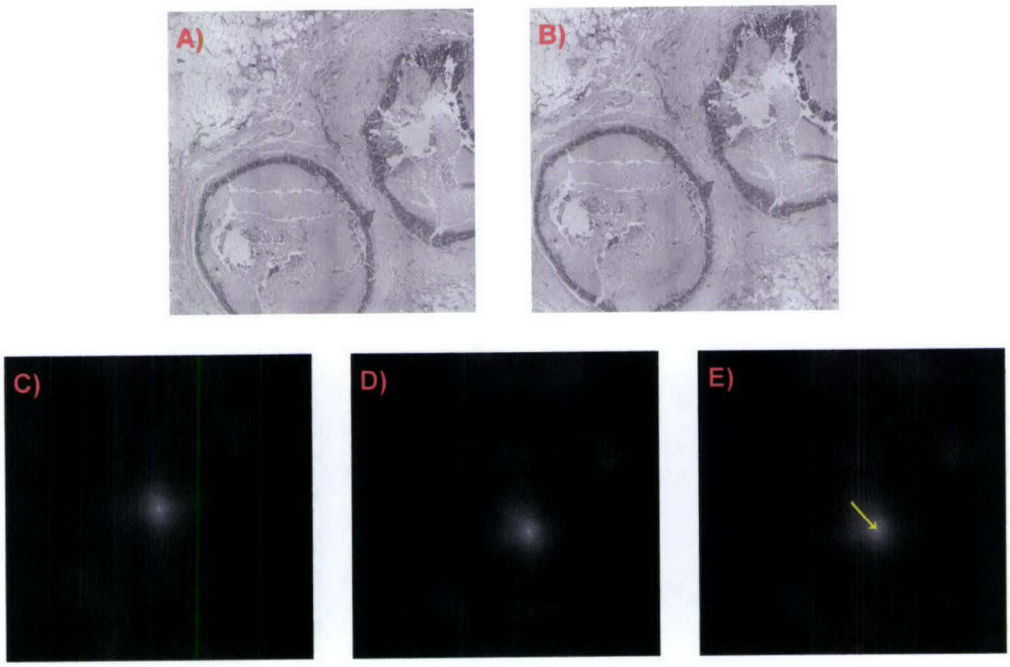


Figure caption: These set of images describe the calculus of the correction vector in the frequency domain. Figures A and B represent exactly the same part of a tissue section image but B is shifted respect to A. Figure C is the modulus image from applying the correlation function to image A (both entries to the function are image A, i.e. this is the calculus of autocorrelation). Figure D shows the correlation result between A and B. Finally, E shows in yellow the correction vector, obtained by subtracting the coordinates of the brightest point of D and the brightest point on.

Since the dimensions of the sub-images are fixed, parts of one or more structures of interest can belong to different sub-images, causing less accurate results than if the entire structures were inside the same sub-images. Therefore the algorithm makes use of two different window sizes: the sub-image size and the correlation area size. The correlation area is the part of the image centered in the sub-image that is used to

calculate the correction vector. Its size must be bigger than the sub-image size in order to avoid the previously mentioned problem. Indeed, this size should equal the estimated maximum error produced during the rigid registration in order to correct the misalignment introduced in that process.

As in the rigid registration algorithm, the local registration algorithm can be applied to the gray scale images or to the binary image containing only the segmented contours. Some problems due to the quality of the images such as improper focusing of areas in the image, different luminosity, or even foreign bodies in the tissue could affect the value of the correlation. Therefore, once the list of correction vectors is obtained, filtering is applied to the vectors to eliminate erroneous vectors. At the end of this process there will be a correction vector and a correlation coefficient for every sub-image. The first filter recalculates the vectors whose correlation coefficient lies under a threshold value, which is obtained using the auto-threshold function over the coefficient list. These correction vectors presenting a low correlation coefficient are substituted by the average of their 3x3 vector neighborhood. Thus, vectors with a very low correlation coefficient will not be taken into account. The second and last filter to be applied to the list of vectors consists in a mean filter, also with a 3x3 kernel, that will allow smoothing the vector field and reducing the effect of the possible noise. The result of applying the vectors to the coordinates before rendering the objects in 3D is more smooth reconstructions, free from errors due to non-linear effects.

Task 3. (Months 6-36) Use #1 & #2 to perform 3-D reconstruction maps of the distribution of ER and PR in the ductal tree and determine whether ER colocalizes with PR in mouse mammary epithelium during critical phases of mammary development.

1. Select the animals (20) to study, taken at different stages of development. Extract, fix and section the tissue as explained in the Methods Section (Months 6-12,4 specimens; Months 12-24, 8 specimens, Months 24-36, 8 specimens)
2. Reconstruct the mammary gland using the software developed according Task 1. (Months 12-24, 5 specimens; Months 24-36, 15 specimens)
3. Revisit areas of interest for high resolution imaging and detection/colocalization of Progesterone and Estrogen receptors) (Months 24-36, 20 sections) Integrate the information (Months 30-36)

All the glands required for this study have been collected. Due to a delay in the development of the software and setting up the immunostaining protocols, we changed the protocol slightly, in that instead of reconstructing and mapping the expression of the receptors in 20 glands from 20 animals (one gland per animal) we used 10 animals. The glands were whole mounted and embedded in paraffin and used to look at the overall expression of the receptors, after being immunostained by color absortion (brightfield). The time points at which the glands were collected represent the main stages of the development of the gland: 6wk (pubertal), 12wk (post pubertal), 18wk (adult), 36wk (menopausal), 48wk (post-menopausal).

The study of ER and PR expression in the mammary gland has been included as Appendix 5, which is also a manuscript (Bibliography J4) that will be soon submitted for publication to a peer-reviewed journal. What follows is a summary of the study, with references to the manuscript.

Briefly, all gland whole mounts where imaged with a digital camera and the percentage of fat pad filled quantified at each time point (See Appendix 5, Results section and Fig. 5). Then the number of end buds visible at the whole mounts was also counted (See Appendix 5, Results section and Fig. 6).

Then, the tissue blocks where sectioned at 5 microns and alternatively stained with H&E or with an antibody to either ER or PR (See Appendix 5, Methods section for details on the protocol used).

All the sections were images at 2.5X using R3D2 (See Appendix 5, Fig. 1). Selected end buds, ducts (small and large) and alveolar structures were segmented on the

images of the H&E sections, and the contours thus obtained were grouped an rendered in 3D (See Appendix 5, Methods sections for details on how the images were processed and reconstructed. See also Fig. 7 for an example of a reconstruction).

Then, selected areas of the reconstructed structures were identified and imaged in the intermediate ER or PR immunostained sections. These color images were taken at 40X magnification. All the cells in ductal epithelial structures were annotated as being ER or PR positive, depending on the particular immunostaining used in the section. Areas corresponding to the same type of structure (i.e. small ducts, large ducts, alveolar structures, end buds) were then analyzed for the number of positive or negative cells for each receptor. (See Appendix 5, Methods section for a more detailed description of the annotation and the analysis).

The analysis (See Appendix 5, Results section and Figs. 9 & 10), shows the evolution of receptor status of the gland, classified by type of structure. It shows dramatic decrease of receptor expression in the transition between puberty and adulthood, after which the expression stabilizes for the rest of the life of the animal. We also showed significant expression differences between small (ducts, alveoli) and large structures (ducts) and similar levels of small structures and end buds. See Appendix 5 for a thorough description of the results and conclusions of the study.

Finally, sections from all the cases were double-immunostained with antibodies to both ER and PR and the number of colocalized cells was counted and represented for all the time points (See Appendix 5, Fig. 11).

KEY RESEARCH ACCOMPLISHMENTS

The main accomplishments achieved from this project are listed below. Some have been already described in the yearly reports.

- We have developed a system that allows semi-automatic 3D reconstruction of tissue samples from fully sectioned tissue blocks. The system allows acquisition of entire tissue sections at low magnification, both in bright field and fluorescence microscope, registration of images of consecutive sections and annotation and 3D rendering of tissue structures (epithelial, endothelial, etc). The system allows revisting of areas of the tissue at higher magnification from both the 3D reconstruction of the tissue as wells as from the low resolution images of the sections. (Bibliography J1 –Appendix 1-, P1, A1, A2, A3, O1, O2, O3, O4)

- To solve the bottlenecks of tissue imaging and reconstruction, we have developed tools that automatically register consecutive sections (Bibliography A3, P3 –Appendix 3-) and for the unmanned segmentation of histological structures using geometrically driven image flows (Bibliography J2 –Appendix 2-, P2)

- We have developed spatial statistics software to quantify pattern of distribution of different cell phenotypes in tissue sections and cells in culture. (Bibliography P4 –Appendix 4-)

- We have used our software to quantify and compare the expression of hormone receptors (ER and PR) in different stages of the mouse mammary gland development . We have determined the levels of expression for different types of tissue structures and the level of colocalization of both receptors. (Bibliography J24 –Appendix 5-, A1,A3,A4)

REPORTABLE OUTCOMES

Manuscripts (published):

- "*A system for combined three-dimensional morphological and molecular analysis of thick tissue samples*" Fernandez-Gonzalez R., Jones A., Garcia-Rodriguez E., Chen P.Y., Idica A., Barcellos-Hoff M.H., Ortiz de Solorzano C. *Microscopy Research and Technique* 59(6):522-530, 2002..
- "*Recent advances in quantitative digital image analysis and applications in Breast Cancer*". Ortiz de Solorzano C., Callahan D.E., Parvin B., Costes S., Barcellos-Hoff, M.H. *Microscopy Research and Technique* 59 (2):119-127, 2002.
- "*A geometric model for image analysis in cytology*" Ortiz de Solorzano C., R. Malladi, Lockett S. In: Geometric methods in bio-medical image processing. Ravikanth Malladi (Ed.). Springer Verlag 2002, pp. 19-42.
- *"Automatic segmentation of histological structures in mammary gland tissue sections". Fernandez-Gonzalez R., Deschamps T., Idica A.K., Malladi R., Ortiz de Solorzano C. Journal of Biomedical Optics 9(3):445-453, 2004..*

Manuscripts (in preparation):

- *"Three-dimensional Histo-pathology of the mammary gland".* Laribi O., Hartland A., Arganda-Carreras I., Fernández-Gonzalez R., Idica A.K., Ortiz de Solorzano C. To be submitted to *The Journal of Mammary Gland Biology and Neoplasia.*
- *"Quantitative Image Analysis in Mammary Gland Biology".* Fernández-González R., Barcellos-Hoff M.H., Ortiz de Solórzano C. Review paper requested for a special *Journal of Mammary Gland Biology and Neoplasia* issue on Quantitative analysis of the mammary gland (to be published in December 2004).

Conference Proceedings:

- "*A system for computer-based reconstruction of 3-dimensional structures from serial tissue sections: an application to the study of normal and neoplastic mammary gland biology*". Fernandez-Gonzalez R., Jones A., Garcia-Rodriguez E., Knowles D., Sudar D., Ortiz de Solorzano C. Proceedings Microscopy and Microanalysis'01. Microscopy and Microanalysis 7, Supplement 2, pp.964-965, 2001
- "*Automatic segmentation of structures in normal and neoplastic mammary gland tissue sections*". Fernandez-Gonzalez R., Deschamps T., Idica A.K.,

Malladi R., Ortiz de Solorzano C., Proceedings of Photonics West 2003, Vol. 4964, 2003

- *"Automatic and segmentation-based registration of serial mammary gland sections".* Arganda-Carreras I., Fernández-Gonzalez R., Ortiz de Solórzano C. Accepted for the IEEE Engineering in Medicine and Biology Society (EMBS) International Conference, San Francisco September 2004.
- *"A tool for the quantitative spatial analysis of mammary gland epithelium".* Fernández-González R., Ortiz de Solórzano C. Accepted for the IEEE Engineering in Medicine and Biology Society (EMBS) International Conference, San Francisco September 2004.

Presentations:

- *"A system for computer-based reconstruction of 3-dimensional structures from serial tissue sections: an application to the study of normal and neoplastic mammary gland biology".* Microscopy and Microanalysis'01, Long Beach, CA August 5^{th}-9^{th}, 2001. Platform presentation.
- *"3D Mammary Histopathology"* Fernandez-Gonzalez R., Idica A. K., Ortiz de Solorzano C. 2003 Mammary Gland Biology Gordon Research Conference . Roger Williams University, Bristol, Rhode Island, June 1-6, 2003.
- *"Automatic Registration of Mammary Gland Section Images"* Arganda-Carreras I., Fernández-González R., Ortiz de Solórzano C. First International Meeting on Applied Physics (APHYS 2003), Badajoz, Spain, October 13^{th}-18^{th}, 2003
- *"Three-dimensional heterogeneity of the mammary gland and breast tumors".* Ortiz de Solórzano C., Gordon Research Conference on Mammary Gland Biology. Il Ciocco, Italy, May 2004.

Informatics:

- As described in the Body of the report and in the Reportable Outcomes sections, we have developed and integrated new methods to acquire and reconstruct and analyze fully sectioned tissue samples. This includes tools for the acquisition, segmentation, and registration of histological sections, as well as cell-level quantitative statistical and spatial analysis of immunostained sections. All these tools are nicely integrated under a single Graphical User Interface. To the best of our knowledge, ours is a

unique piece of software, since there is no commercial or reported system that can perform all these tasks under a single software platform.

Funding obtained:

- *Segmentation of Mammary Gland Ductal Structure Using Geometric Methods.* P.I.'s Malladi R. and Ortiz de Solorzano C. Granted by the LBNL Laboratory Directed Research and Development Program (LDRD), in the Strategic-Computational Sub-Program. Period Oct 2001- Sept 2004
- *Characterization of Adult Stem Cell Involvement in Mammary Gland Development.* PI: Dr. Carlos Ortiz de Solorzano Funded by: LBNL Laboratory Directed Research and Development Program (LDRD). Period Oct 2002-Sept 2004
- *Three-dimensional Modeling of breast cancer progression.* PI: Dr. Carlos Ortiz de Solorzano. Funded by: University of California, Breast Cancer Research Program Grant Number – 8WB-0150
- *"Characterization of label-retaining cells and their niche in the mouse mammary gland"* DOD Breast Cancer Research Program Predoctoral Fellowship. PI. Rodrigo Fernández-González.
- NASA NSCOR Program Project. PI. Mary Helen Barcellos-Hoff, Ph.D.

Funding applied:

- "*Biological Basis and Functional Phenotypes of Breast Density*". NIH Program Project. PI: Thea Tlsty, Ph.D.
- "*An automated system for three-dimensional Histopathology*". NIH R21/R33 Phase Innovator Award, for the PAR: "Technology developments for Biomedical Applications" PI. Carlos Ortiz de Solórzano, Ph.D.

Employment or Research:

- Due to the successful performance of the PI as a Scientist during the first two years of the project, in 2002 he was promoted to a Staff Scientist Position at the Life Sciences Division, Lawrence Berkeley National Laboratory of the University of California.

- This grant has partially supported Mr. Rodrigo Fernández-González, a Ph.D. candidate in the joint UC Berkeley-UC San Francisco Program in Bioengineering. Rodrigo continues working with me part time as a Graduate Student Research Assistant. On January 2003 he was granted a DOD-BCRP predoctoral fellowship *"Characterization of label-retaining cells and their niche in the mouse mammary gland* " that will support him until the end of his graduate work.
- Half way through the reporting period (in January 2002), Dr. Umesh Adiga, a Ph.D. in Computer Sciences, joined my lab as a postdoctoral fellow to work on the image analysis required for the automatic segmentation of nuclei and FISH signals, as well to other image analysis and processing tools required for this project. He left the laboratory in January 2003. Unfortunately, his performance was lower than expected, based on his previous work and references.
- In October 2002 Dr. Ouahiba Laribi, a Ph.D. in Molecular and Cell Biology joined the lab. She has been a phenomenal help in the last two years.
- Mr. Adam Idica, an Integrated Biology undergraduate student at UC Berkeley worked in this project for three years. He graduated from UC Berkeley in May 2003 and has continued working in our lab as Technical Assistant until July 2004, when he continued his studies towards a Medical Degree..
- Two other undergrad students, Ms. Abbey Hartland and Mr. Shamroze Khan, from Shasta College (Redding, CA) and UC Berkeley, have worked in my lab as undergraduate technical assistants and have greatly contributed to produce the quantitative results of this project.

CONCLUSIONS

In summary, during the administrative length of the project, we have developed the computer and microscopy platform that we proposed to develop. The developments were slower than expected due to the realization of the need for automating some of the time consuming tasks, namely the registration and annotation of the images of the sections. The developments have been or are in the process of being published in peer-reviewed journals and have been successfully presented in international conferences. We believe that this advanced computerized microscopy platform, developed thanks to the funding support of this grant, will be of use in many future studies that required looking at molecular events at cellular level within the tissue context where they occur.

Due to the above mentioned delay in the technical developments, added to other personnel issues and a delay in setting up the immunostained protocols, the application of the new system to study the expression of the hormone receptors in the entire gland, at different time points, we analyzed 10 of the 20 samples initially proposed. No loss of potential information was caused by this reduction, since the samples covered to the three main stages of development (puberty, adulthood, menopause), and a very high number of cells were used from each animal. The result is a description of the expression of the receptors in those stages of development as well as of the differences in expression between different histological structures.

BIBLIOGRAPHY

Papers in Peer Reviewed Journals:

J1. A system for combined three-dimensional morphological and molecular analysis of thick tissue samples. Fernandez-Gonzalez R., Jones A., Garcia-Rodriguez E., Chen P.Y., Idica A., Barcellos-Hoff M.H., Ortiz de Solorzano C. *Microscopy Research and Technique* 59(6):522-530, 2002. **(Appendix 1)**

J2. Applications of quantitative digital image analysis to breast cancer research. Ortiz de Solorzano C., Costes S., Callahan D.E., Parvin B., Barcellos-Hoff M.H. *Microscopy Research and Technique* 59(2):119-127, 2002

J3. Automatic segmentation of histological structures in mammary gland tissue sections. Fernandez-Gonzalez R., Deschamps T., Idica A.K., Malladi R., Ortiz de Solorzano C. *Journal of Biomedical Optics* 9(3):445-453, 2004. **(Appendix 2)**

J4. Three-dimensional Histo-pathology of the mammary gland. Laribi O., Hartland A., Arganda-Carreras I., Fernández-Gonzalez R., Idica A.K., Ortiz de Solorzano C. To be submitted to *The Journal of Mammary Gland Biology and Neoplasia.* **(Appendix 5)**

J5. Quantitative Image Analysis in Mammary Gland Biology. Fernández-González R., Barcellos-Hoff M.H., Ortiz de Solórzano C. Review paper requested for a special *Journal of Mammary Gland Biology and Neoplasia* issue on Quantitative analysis of the mammary gland. (to be published in December 2004)

Book chapters:

B1. A geometric model for image analysis in cytology. Ortiz de Solorzano C., R. Malladi, Lockett S. In: Geometric methods in bio-medical image processing. Ravikanth Malladi (Ed.). Springer Verlag 2002, pp. 19-42.

Conference Proceedings:

P1. A system for computer-based reconstruction of 3-dimensional structures from serial tissue sections: an application to the study of normal and neoplastic mammary gland biology. Fernandez-Gonzalez R., Jones A., Garcia-Rodriguez E., Knowles D., Sudar D., Ortiz de Solorzano C. Proceedings Microscopy and Microanalysis'01. Microscopy and Microanalysis 7, Supplement 2, pp.964-965, 2001

P2. Automatic segmentation of structures in normal and neoplastic mammary gland tissue sections. Fernandez-Gonzalez R., Deschamps T., Idica A.K., Malladi R., Ortiz de Solorzano C., Proceedings of Photonics West 2003, Vol. 4964, 2003

P3. Automatic and segmentation-based registration of serial mammary gland sections Arganda-Carreras I., Fernández-Gonzalez R. Ortiz de Solórzano C. Accepted for the IEEE Engineering in Medicine and Biology Society (EMBS) International Conference, San Francisco September 2004. **(Appendix 3)**

P4. A tool for the quantitative spatial analysis of mammary gland epithelium. Fernández-González R., Ortiz de Solórzano C. Accepted for the IEEE Engineering in Medicine and Biology Society (EMBS) International Conference, San Francisco September 2004. **(Appendix 4)**

Abstracts:

A1. Three-dimensional Characterization of Ductal Carcinoma In Situ of the Breast. Carlos Ortiz de Solórzano, Rodrigo Fernández-González, Adam Idica, Fred Waldman, Joe W. Gray. Era of Hope Meeting. Orlando, FL. September 25-28, 2000.

A2. 3D Mammary Histopathology. Fernandez-Gonzalez R., Idica A. K., Ortiz de Solorzano C. 2003 Mammary Gland Biology Gordon Research Conference. Roger Williams University, Bristol, Rhode Island, June 1-6, 2003.

A3. Automatic Registration of Mammary Gland Section Images. Arganda-Carreras I., Fernández-González R., Ortiz de Solórzano C. First International Meeting on Applied Physics (APHYS 2003), Badajoz, Spain, October 13th-18th, 2003

A4. Three-dimensional heterogeneity of the mammary gland and breast tumors. Ortiz de Solórzano C., Gordon Research Conference on Mammary Gland Biology. Il Ciocco, Italy, May 2004.

Online references:

O1. Laser Focus World:
http://lfw.pennnet.com/Articles/Article_Display.cfm?Section=ARCHI&ARTICLE_ID=177510&VERSION_NUM=1&p=12

O2. Cipherwar Science new service: http://www.cipherwar.com/lists/lbnl/msg00025.html

O3. DailyCalifornian Science and Technology:
http://www.dailycal.org/article.asp?id=11347&ref=search

O4. Lawrence Berkeley National Laboratory New Release:
http://www.lbl.gov/Science-Articles/Archive/LSD-tracking-cancer-3D.html

PERSONNEL

- Carlos Ortiz de Solórzano, Ph.D. Principal Investigator
- Umesh Adiga, Ph.D. Postdoctoral Fellow (from 01/02 to 01/03)
- Ouahiba Laribi, Ph.D. Postdoctoral Fellow. (from 11/02 to 07/04)
- Rodrigo Fernández-González, Graduate Student Research Assistant (from 10/00)
- Ignacio Arganda-Carreras, Graduate Student Research Assistant (from 10/02)
- Adam Idica, Technical Assistant (from 10/01)
- Shamroze Khan, Undergraduate Technical Assistant (from 05/03)
- Abbey Hartland, Undergraduate Technical Assistant (from 06/03)

MICROSCOPY RESEARCH AND TECHNIQUE 59:522–530 (2002)

System for Combined Three-Dimensional Morphological and Molecular Analysis of Thick Tissue Specimens

RODRIGO FERNANDEZ-GONZALEZ, ARTHUR JONES, ENRIQUE GARCIA-RODRIGUEZ, PING YUAN CHEN, ADAM IDICA, STEPHEN J. LOCKETT, MARY HELEN BARCELLOS-HOFF, AND CARLOS ORTIZ-DE-SOLORZANO*

Life Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720

KEY WORDS computer assisted microscopy; 3D reconstruction; assisted histopathology; JAVA

ABSTRACT We present a new system for simultaneous morphological and molecular analysis of thick tissue samples. The system is composed of a computer-assisted microscope and a JAVA-based image display, analysis, and visualization program that allows acquisition, annotation, meaningful storage, three-dimensional reconstruction, and analysis of structures of interest in thick sectioned tissue specimens. We describe the system in detail and illustrate its use by imaging, reconstructing, and analyzing two complete tissue blocks that were differently processed and stained. One block was obtained from a ductal carcinoma in situ (DCIS) lumpectomy specimen and stained alternatively with Hematoxilyn and Eosin (H&E), and with a counterstain and fluorescence in situ hybridization (FISH) to the ERB-B2 gene. The second block contained a fully sectioned mammary gland of a mouse, stained for histology with H&E. We show how the system greatly reduces the amount of interaction required for the acquisition and analysis and is, therefore, suitable for studies that require morphologically driven, wide-scale (e.g., whole gland) analysis of complex tissue samples or cultures. *Microsc. Res. Tech. 59:522–530, 2002.* Published 2002 Wiley-Liss, Inc.†

INTRODUCTION

Understanding complex biological systems requires tissue-level integration of information from multiple sources such as molecular, physiological, anatomical, and so on. However, none of the existing analytical methods in biology provides the required level of integration, in that they either do not account for intercellular variation (RFLP, Southern blots, microarray technologies, etc.), or they do it at the expense of tissue integrity (e.g., flow cytometry).

Image-based cytometry (IC) can provide molecular or genetic information (e.g., by using fluorescence in situ hybridization [FISH] or immunohistochemistry [IHQ]), in fixed cells within their native morphological tissue context. Volumetric, 3D morphological information can be obtained using confocal laser scanning microscopy (CLSM). However, only relatively thin tissue sections (<100 μm) can be studied, due to light scattering and refractive index mismatch problems that occur when imaging deeper in the tissue, and because of practical limitations of effectively staining thicker sections by IHQ or FISH.

When analysis and integration of molecular and morphological information at high resolution is aimed on a bigger scale (e.g., tissue or a small gland), the only existing approach requires sectioning the tissue, followed by both histological and molecular staining of consecutive tissue sections. The analysis is normally performed by visual inspection of the sections under the microscope. This approach greatly limits the extent and accuracy of the analysis, due to the difficulty that our visual system finds when composing (extrapolating) meaningful 3D information from a series of 2D sections. Since only a few colors (2–4) can be discriminated, both in bright field and fluorescence microscopy, along with other practical problems related to multicolor IHQ or (F)ISH, we have to do the histological and the molecular staining on different, alternative sections. This further complicates the visual analysis and integration of molecular and morphological information.

To overcome these problems, we have developed a three-dimensional microscopy system that integrates computer analysis and visualization tools. These tools automate or greatly reduce the amount of interaction required for the acquisition, reconstruction, and morphologically directed analysis of thick tissue samples. Our system can be used to reconstruct tissue structures and to quantitatively measure the presence and spatial distribution of different molecular elements (e.g., genes, RNAs, proteins) in their intact cellular environment. This tool is currently being used to study breast cancer, where heterogeneity and three-dimensionality are at the very base of both disease initiation and clonal progression.

Our system encapsulates a three-dimensional visualization system and an image analysis system. The application was developed using a distributed architecture (client-server model), and Java for writing the graphical user interface (GUI) so that it can run remotely on any computer platform. The system allows

Grant sponsor: Department of Defense Breast Cancer Research Program; Grant numbers: DAMD17-00-1-0306, DAMD17-00-1-0227.

*Correspondence to: Carlos Ortiz de Solorzano, Ph.D., Ms 84-171, Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, CA 94720. E-mail: CODeSolorzano@lbl.gov

Received 20 December 2001; accepted in revised form 27 June 2002

DOI 10.1002/jemt.10233

Published online in Wiley InterScience (www.interscience.wiley.com).

acquiring and registering low magnification (e.g., 1 pixel = 5 μm) conventional (bright field or fluorescence) images of entire tissue sections. It can also create a virtual 3D reconstruction of the tissue structures, from which new areas of interest can be revisited or reacquired at high resolution (e.g., 1 pixel = 0.5 μm) and automatically analyzed.

Although there are existing commercial and noncommercial software packages that can separately perform some of those functions, the integration of all of them (multiresolution, multicolor acquisition of multifield fluorescence microscopy images; reconstruction of structures on interest in 3D; molecular and morphological 3D analyses; content-based image and data storage and retrieval system through the use of "Cases"; or a series of consecutive images and data belonging to a given tissue) on a single platform makes our system a very powerful tool.

A comparison to similar systems shows that many of them offer a set of independent programs running under different interfaces (*IMOD*, Boulder Laboratory for 3-Dimensional Fine Structure), but not a common platform that integrates all of them. Most of these systems are not designed to run on multiplatform environments (VIDA, University of Iowa; Trace, Boston University), some others have special hardware requirements for real time rendering (VoxBlast, Vaytek Inc.) and many lack a proper image management system (Imaris, Bitplane AG). In this review, we describe our system and we illustrate its use by presenting the reconstruction of a mouse mammary gland and a tumor biopsy of a patient with ductal carcinoma in situ (DCIS) of the breast.

MATERIALS AND METHODS

System Description

The system is controlled by a client-server application, as shown in Figure 1. The server is a C language application that runs on a computer (Dell Inspiron, running Solaris 7 for Intel) connected to an Axioplan (Zeiss Inc., Germany) microscope. The server actuates all the moving parts of the microscope: motorized scanning stage, excitation filter wheel, and arc-lamp blocking shutter (Ludl Electronic Products Ltd., Hawthorne, NY). It controls the CCD MicroImager camera (Xillix Techonologies Corp., Richmond, British Columbia, Canada) as well. The server can perform basic operations, such as acquiring and storing images, setting the exposure time of the CCD, moving the stage, and operating the filter wheel. In addition, it has been programmed to offer more complex functions, such as automatically focusing the microscope or acquiring multiple field-of-view images. To do this, the server receives each order and divides it into a set of simple actions. For instance, to acquire a multiple field-of-view image, the server asks for the coordinates of the vertices of the area to be acquired and then automatically performs the required sequence of stage movements and camera acquisitions. The output is a mosaic-like image of the area. The server can do multi-color acquisition in fluorescence and bright field, by performing consecutive acquisition using different excitation filters and multi-band emission filters.

Description of the Client Application. The client (R3D2) is connected to the server through UNIX sockets, which are the standard for Internet-based communications. It can send requests to the server from any computer connected to the Internet. To obtain the maximum benefit from this, R3D2 has been written in JAVA (v.1.2), so that it can be executed on many different computer platforms.

Figure 2 shows the R3D2's complete Graphical User Interface (GUI). The interface is divided in two distinguishable parts. One (rightmost vertical panel in Fig. 2) provides connections to the server and allows the user to request its services through a user-friendly interface. The available actions can be classified in two groups.

Basic control operations. These include all the simple, atomic, operations provided by the server, such as setting the objective lens, changing the excitation filter (fluorescence), setting the exposure time of the camera, moving the stage to the absolute origin of coordinates relative to which all measurements are taken, opening/closing the arc lamp blocking shutter, and acquiring a single image using the current microscope settings. R3D2 receives the image from the server and displays it, both complete (zoomed out), as well as partially (in its original resolution). Only a small part of the full-resolution image can be displayed at full resolution. The zoomed area can be interactively selected by moving a window on the complete version of the image (Fig. 3). Images can be *saved* both in ICS (Dean et al., 1990) and JPEG format. When images are saved as ICS, all the acquisition parameters (objective, filter, location of the image on the slide are, etc.) are stored in the ICS header file. JPEG format, compressed or not, can be used as an alternative format when the user does not plan any future analysis of the images, and images are stored for exchange, document creation, or web publishing.

Complex operations. These operations combine multiple atomic operations to provide the following functionality.

- Autofocus: Automatically focuses the microscope by taking a series of images at different positions in the Z axis (step size = 0.50 μm for low magnification images, 0.25 μ for high magnification images) and determining the best-focused image of the series. Blur, due to out-of-focus light, reduces image contrast, which can be detected using several functions. Based on several comparisons described in the literature (Firestone et al., 1991; Groen et al., 1985; Santos et al., 1997), we selected an autocorrelation-based function introduced by Vollath (1987).
- Scan: Acquires multiple field of view images. The system displays a dialogue-panel where the user can specify the filter(s) to be used (in fluorescence microscopy), exposure time(s) and the limits of the area to acquire. The limits can be defined by its coordinates (when known) or manually, by moving the microscope to the upper, lower, rightmost, and leftmost points of the area.
- Revisit Point: When the user clicks on a point of the image of a previously acquired complete or partial tissue section, the server moves the stage to that location on the slide and takes an image using the current values of objective, filter, and exposure time.
- Revisit Area: When using this option, the user is

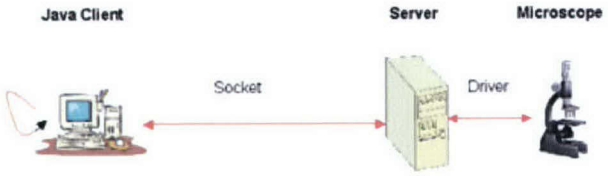


Fig. 1. Description of the client-server architecture of R3D2. The server runs on a computer with a microscope attached, and provides access to the microscope functions. The client is a JAVA application, which can run on any computer (no particular OS required) connected to the server through the Internet Protocol (IP). The communication between client and server uses UNIX sockets. The client provides user-friendly access to all the microscope functions offered by the server, and allows handling of sets of related images (Cases) for storage, annotation, and 3D reconstruction of structures of interest. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

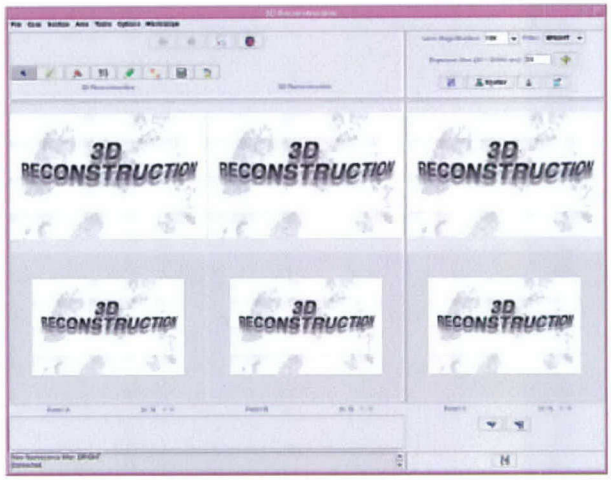


Fig. 2. R3D2's Graphical User Interface. The GUI is divided in two main parts. **Left:** Display consecutive sections of a Case. It also provides the user with tools for registering sections, annotating the images, connecting structures between sections (e.g. mammary ducts, tumor volumes), and reconstructing the annotated images in 3D. **Right:** Access to the microscope related functions offered by the Server. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

asked to draw a rectangle on one image of a previously acquired section. The selected area will be then acquired with the microscope settings provided by the user. Multicolor area acquisition is an option as it is for scanning complete sections.

The second part of the interface (two left vertical panels in Fig. 2) expands the system capabilities, by allowing creating and handling sets of related images, which we call Cases. A Case is a sequence of low-magnification images of complete tissue sections taken from a tissue block, along with all the areas re-visited on them at higher resolution and with different filters, plus the results of analysis performed on them if any. The image files that make up a Case are specially labeled for convenience. The user can:

- Annotate the Cases, by marking and/or delineating structures of interest and linking them within and between consecutive sections. The user can add textual annotations (Text), ductal structure identifiers (with a unique number that identifies them within the section, Duct) and forms that delineate irregularly shaped ducts or other structures (Shapes). In addition to this, ductal marks can be connected within the same section or in different sections (Connected Ducts), and corresponding shapes in different sections can be grouped (Groups).
- Register acquired sections. Before reconstructing a Case in 3D, all its sections must be registered to ensure proper alignment of the elements that will be later reconstructed. For that, we calculate the Rigid-Body Transform that provides the optimum rotation and translation between each pair of sections. The Transform is calculated from three pairs of points interactively marked by the user on each pair of images to be registered. Once the points have been marked, the software calculates the rotation and translation (θ, t_x, t_y) needed to minimize the sum of the squared distances between all three pairs of corresponding points, thus aligning both images. The results are stored in the second image. This method is very accurate when the pairs of points are spread all along the sections. Reasons for small errors are imprecise mouse interaction, stretching and/or compression of the tissue, and the fact that some structures used to select pairs of corresponding points might not be perpendicular to the sections.
- Reconstruct Cases. Our system reconstructs the tissue structures by rendering the user annotations in 3D. Besides the obvious advantage of volumetric tissue structure visualization, the 3D rendering is linked to the microscope for revisiting, and to the original images and their analysis for display of the images and analysis results, as will be described later. The 3D reconstruction part of the software has been developed using Java3D (v.1.2.1.) This is an application-programming interface (API) for 3D in Java. After asking the user for the range of sections to render, the system converts the coordinates of all the markings in that range of sections from two-dimensional to three-dimensional values. The section number and thickness, along with the distance between sections, determines the depth-coordinate. Then a 3D scene is built using several geometric shapes to represent the different markings. Duct markings are rendered as spheres and connected with lines within and between sections. Contour Shapes are rendered as volumes after applying a refined Delaunay triangulation, using the Nuages reconstruction software (INRIA, France, http://www-sop.inria.fr/prisme/) (Boissonnat et Geiger, 1993). The 3D rendition of the Case is displayed in a new window where the mouse can rotate the scene, zoom in or out or translate the scene in the X and/or Y directions. The 3D window includes a tool bar with options to select elements. By just clicking on one element of the 3D scene, the user can get information about it (location, size, etc.), load the image(s) that contain that element, images are displayed in a new image panel, or move the microscope to its location

on the slide for re-imaging. Volume selection is handled by JAVA 3D. The selection is performed by tracing a "virtual ray" from the user's point of view (defined when rendering the Case, normally at a point corresponding approximately to the position of the user's eye) and the point where the user clicks on the screen. The selected volume will be the first object intersected by the ray within the 3D scene. The user can also hide or show all the different elements of the scene, reset it to the default view and change the scale in any of the three dimensions.

- Analyze Cases. All areas selected based on the 3D morphological reconstruction, can be batch processed upon a user request. This way, only those areas selected based on a particular morphological feature are analyzed, and not all the tissue sections, thus reducing the amount of work required. The analysis is done by streamlining the selected images to a new process running custom-made image analysis routines built on a commercial image processing software (Scilimage, TNO, The Netherlands). At the moment, the image analysis routines can segment counterstained nuclei and detect and quantify FISH probes or punctuate-patterned expressed proteins. The image analysis algorithms for nuclei and signal segmentation have been described elsewhere (Malpica et al., 1997, Ortiz de Solorzano et al., 1998, Malpica et al., 1999). The results of the analysis can be displayed from the 3D rendering window, and global measurements can be performed after selecting the volumes.

An important feature of R3D2 is that all Case-handling and marking functions can be used in parallel to the functions that request microscope actions. Therefore, acquiring a new section can be done in parallel to any other Case related function (e.g., registering already acquired sections or annotating the images). Our implementation of this feature uses Solaris threads. Threads permit executing multiple parallel copies of a program without multiplying resource use. Each thread shares memory and other resources with other threads. R3D2 runs on a main thread and when a microscope-based operation is selected, it launches a new thread that runs on the same memory space as the main one. This scheme guarantees that, in the case of a microscope failure or a socket error, the system will not die abruptly, as only the thread working on the microscope will be affected.

RESULTS

We will now illustrate the use of our system by showing how two different tissue blocks were imaged and reconstructed. The first one (HB) is a tissue block from the mammary gland of a patient with ductal carcinoma in situ of the breast (DCIS). The second block (MB) is a normal mammary gland of a nuliparus mouse.

Tissue Source

HB. The tissue was part of a breast lumpectomy specimen. After surgery, the specimen was fixed in an alcoholic-formalin solution, and embedded in paraffin per routine at the California Pacific Medical Center in 1981. The tissue was originally staged as a T1N0M0

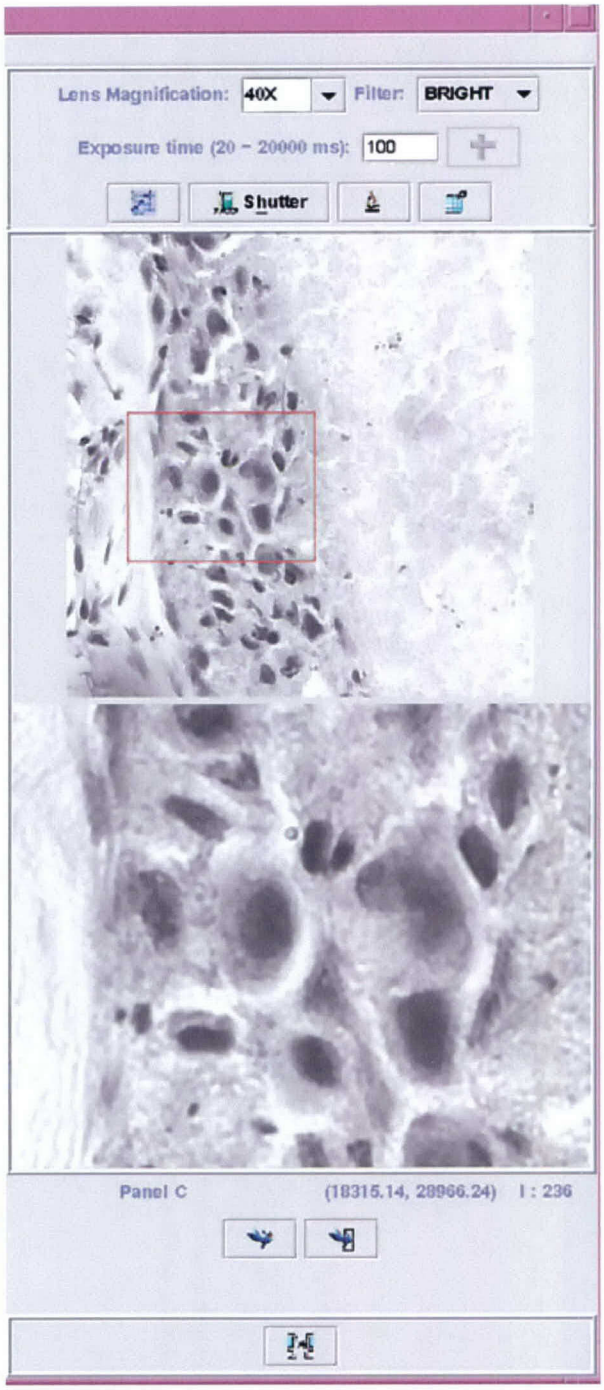


Fig. 3. Single image acquisition. To acquire a single image, the user clicks on the "Acquire" button, after choosing the appropriate microscope settings: objective lens, excitation filter (in fluorescence), and exposure time of the CCD camera. The "Autofocus" option automatically finds the correct focus plane of the microscope before acquiring an image. **Top:** The entire image (reduced to fit in the window). **Bottom:** Contains a zoomed version of the part of the original image selected by the rectangle on the top window. The rectangle can be manually moved to look at different parts of the image at its original, full, resolution. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

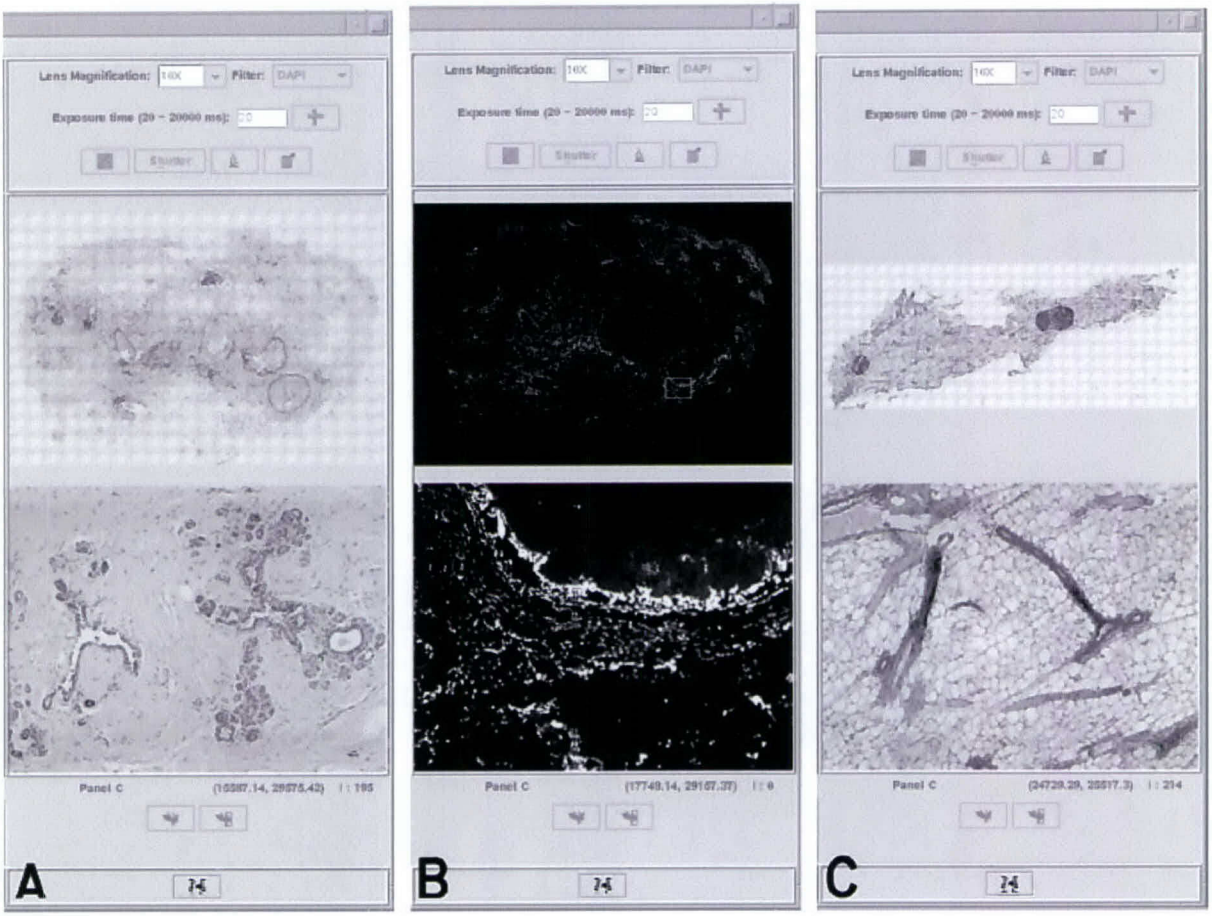


Fig. 4. Tissue acquisition. Images of full sections of a tissue block of ductal carcinoma in situ of the breast (**A,B**) and of a mouse mammary gland (**C**). Both blocks were paraffin embedded, sectioned, and stained for Histology (H&E, images A and C) or counterstained with DAPI (B). Microscope focusing and image acquisition is completely automated, as described in the text. In the individual images, the**top** contains the entire section, the **bottom** contains a zoomed sub-area of it.

(Stage 1). Several blocks of approximately 1" × 1" × 3 mm were taken from the tissue. The block we used contained DCIS and benign ducts; there was no invasive tumor present. The tumor was positive for ERBB2 by immunohistochemistry performed on a 4-μm section taken from the top of the block.

MB. Female BALB/c mice were obtained from Simonson (Gilroy, CA) and housed 4 per cage with chow and water ad libitum in a temperature- and light-controlled facility. Carbon dioxide inhalation was used to kill the animals in accordance with the Association for Assessment and Accreditation of Laboratory Animal Care guidelines and institutional review and approval. The 4th inguinal glands were removed for histology and whole mounts at 10 weeks. The tissue was formalin-fixed and paraffin-embedded.

Tissue Preparation

HB. The human block was sectioned at 4 μm thickness. Every 5th section was collected onto a plus (treated) slide, for a total of 66 sections. From this set, every other section was collected and stained with H&E (33 sections, 40 μm apart). The rest of collected sections were sent to the UCSF Cytogenetics core for FISH (33 sections, 40 μm apart), with a probe for the ERBB2 gene, which is amplified in 30% of carcinomas of the breast. The ERBB2 probe was RMC17P077, a P1 probe. The DNA was extracted and labeled by Nick translation with red CY3 dUTP fluorochrome. In summary, the odd-numbered sections were H&E stained and the even-numbered sections were DAPI counterstained and processed for FISH.

MB. The mouse gland was paraffin-embedded, H&E stained, and sectioned at 4 μm. We starting collecting every fifth section (20 μm apart); 80 μm into the tissue, we switched and collected all remaining sections (4 μm apart), for a total of 24 sections. The collected sections were all placed on glass slides for microscopy and imaging.

Imaging of Tissue Sections

HB. Full tissue sections were imaged in bright field (odd-numbered sections) (Fig. 4A), or fluorescence (even-numbered sections) (Fig. 4B), using R3D2's Scan

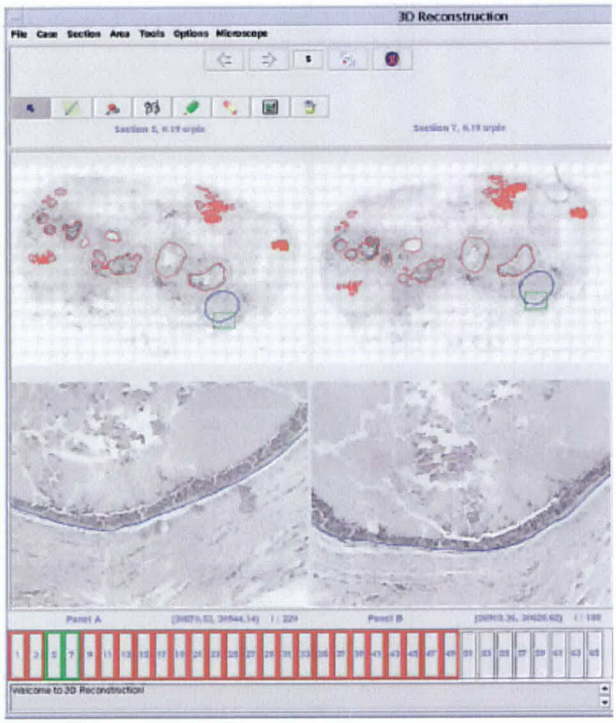


Fig. 5. Tissue annotation (human tissue). Two consecutive H&E stained sections of a human DCIS Case. Only H&E sections are now being shown. The images show manual annotations, namely ducts and tumor masses. Ducts are connected within and between sections. Tumor contours are drawn and also connected between consecutive sections. Connections between sections are shown by changing line color of all connected components every time one of them is visited (selected or just traveled over by the mouse pointer). [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

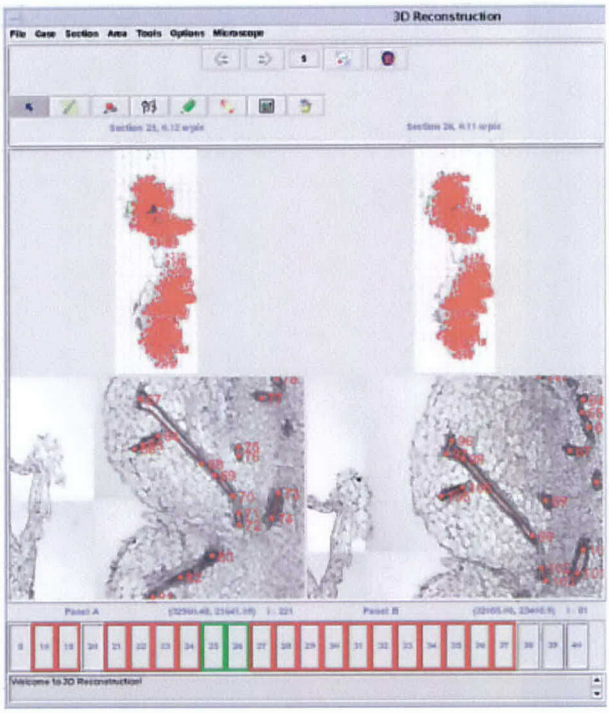


Fig. 6. Tissue annotation (mouse tissue). Two consecutive H&E stained sections of the mouse mammary gland used to test the system. Ducts are connected within and between sections. Lymph node countours are drawn and also connected between consecutive sections. Connections between sections are shown by changing line color of all connected components every time one of them is visited (selected or just traveled over by the mouse pointer). [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

option. Fluorescent sections were imaged using a single-band filter block (360 nm excitation) for DAPI.

MB. All sections were imaged in bright field (Fig. 4C).

In both blocks, sections were imaged at 10× with a Fluar (0.5 n.a.) objective lens (Zeiss, Wetzlar, Germany). In order to optimize memory use while keeping enough resolution for histology, images were reduced by a factor of 4 in both X and Y directions, which gave us an effective 2.5× magnification, i.e., a sixteen-fold reduction in image size. After image compression, the average image size of the sections was 25 Mbytes in the human block and 10 Mbytes in the mouse. The compression was necessary to speed up image transmission and display.

Creation of Cases

Two Cases were created, initially empty. The acquired sections were added with a number equal to its section number within the tissue. All sections were manually registered as previously described to ensure proper alignment of the to-be-done annotations. Each section was registered to its previous section, thus aligning all sections with the first H&E section of the block.

Tissue Annotation

HB. All H&E sections were manually annotated. We marked the centers of the ducts by placing a circular mark (R3D2's Duct tool) in the lumen, when the duct was perpendicular to the image plane, or with a line, when the duct was sectioned longitudinally. Then we connected the markings within and between consecutive sections using the Connect Ducts option. We also delineated tumor areas using R3D2's Shapes option and grouped their consecutive sections using the Group option. (Fig. 5).

MB. The same procedure was followed for the mouse except for the Shapes, which were not used to delineate tumor areas, but the lymph nodes (Fig. 6).

3D Reconstruction

The reconstruction of the Cases, based on the H&E sections, is shown in Figures 7(HB) and 8 (MB). All the user markings are interactive, in that by clicking on them the user can (1) obtain positional information, (2) load the part of the original image section(s) containing that marking, or (3) revisit the selected marking under the microscope. For the latter, the microscope automatically moves to the position of the marking in the tissue, provided that the right slide is on the microscope. In situ tumors and lymph nodes were rendered as volumes in 3D, and ducts as lines.

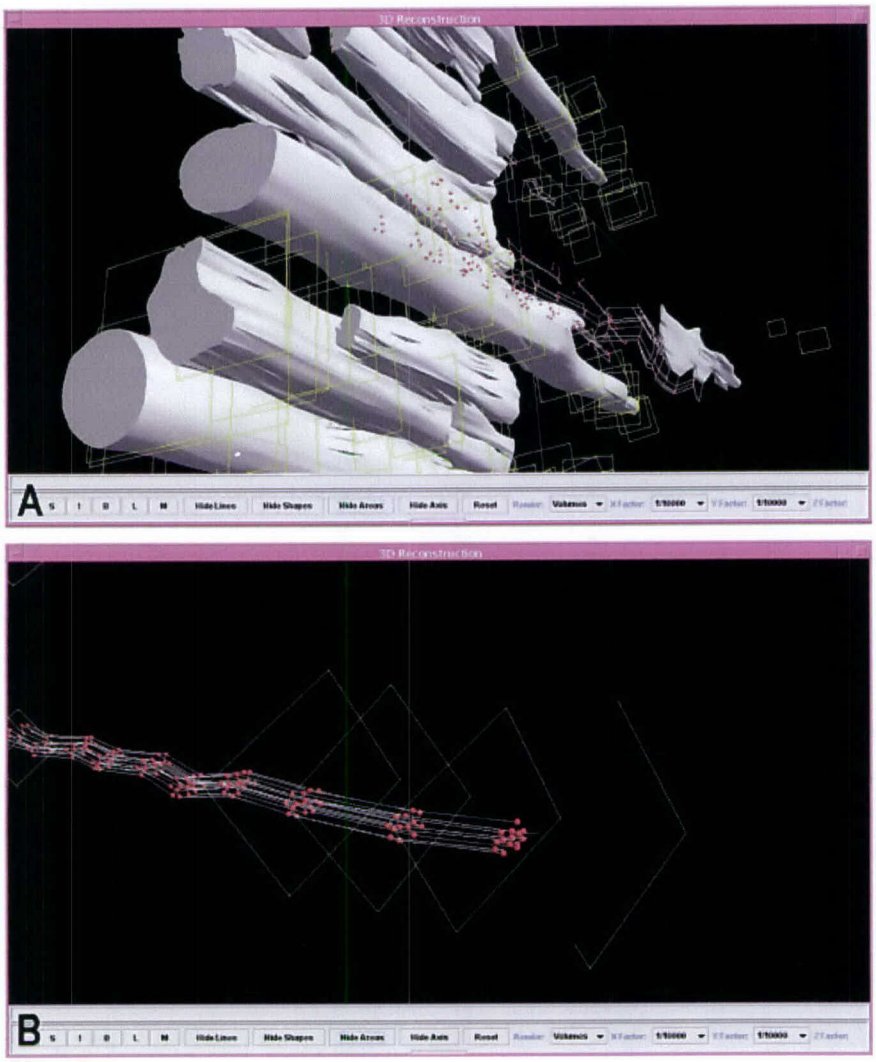


Fig. 7. Tissue reconstruction (human case). 3D reconstruction of the manually annotated DCIS Case. **A:** Tumor volumes have been surface-rendered to show their three-dimensional shapes. Ducts are identified by spheres and connected by lines within and between sections. Yellow rectangles identify areas that were re-acquired at higher magnification. **B:** Close up view of a set of connected markings corresponding to the same duct. The 3D reconstruction is fully interactive. It can be rotated, zoomed, and all the elements can be selected to retrieve information about the element, display the original image that contains the element, or automatically move the microscope to the selected point.

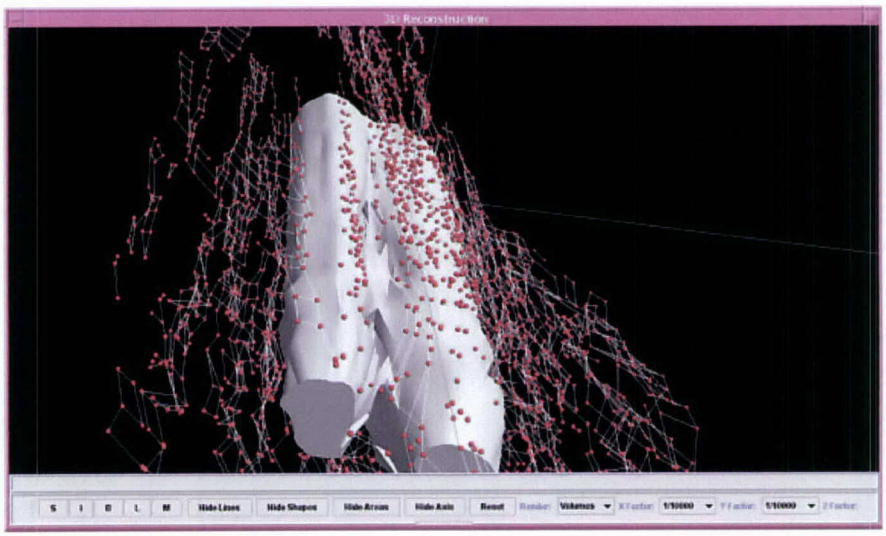


Fig. 8. Tissue reconstruction (mouse Case). 3D reconstruction of the H&E stained, manually annotated mouse Case used to test our system. Lymph node volumes have been surface-rendered to show their three-dimensional shapes. Ducts are identified by spheres and connected by lines within and between sections.

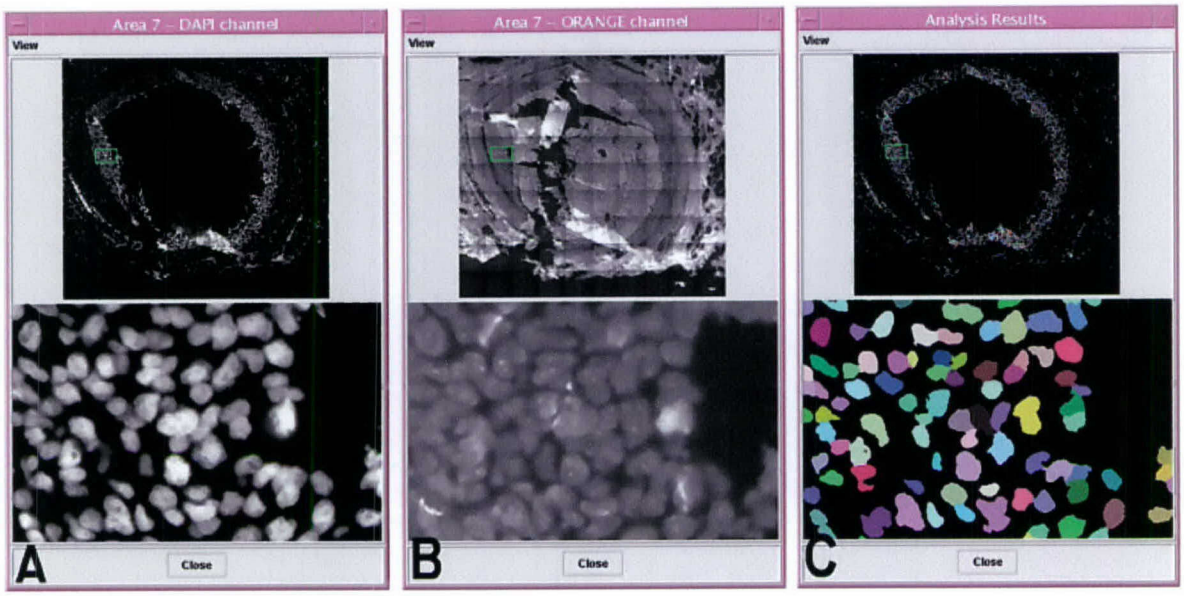


Fig. 9. Revisiting and analysis of areas of interest. Revisited area of the Human DCIS tissue block used to test the system. **A:** DAPI image of counterstained nuclei; **B:** CY3 images of FISH with a probe to the erb-b2 gene. Areas can be acquired and loaded from the images or the sections they belong to or directly from the 3D reconstruction of the tissue. **C:** Results of the segmentation of nuclei based on the counterstained channel. Each segmented nucleus is colored differently. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

Revisit of Areas of Interest

HB. Several areas were revisited at high-resolution (40×), with a Plan-Neofluar (1.3 na) oil immersion lens (Zeiss, Wetzlar, Germany). Areas were manually selected having either morphologically normal or abnormal (DCIS) areas. Areas were double-scanned using a Pinkel filter set (Chroma Technologies, Brattleboro, VT), by consecutively imaging while exciting the sample at 360 (DAPI emission from the nuclei) and 572 nm (CY3 emission from ERBB2 gene). All areas were acquired at full resolution and then compressed by a factor of 2 in both X and Y directions for an effective 20× magnification. Areas were manually selected by drawing rectangles on the low-resolution images of the whole fluorescent sections, although they could have been selected from the 3D reconstruction as well. To ensure proper alignment between the areas and the sections, the images of the sections were previously aligned with the microscope slide by calculating the shift between the current location of the slide and its location when the image of the section was originally acquired. This is done by calculating the rigid body transform between three points manually selected in the image of the section and the corresponding points under the microscope. Figure 7 shows the 3D reconstruction of the human Case, incorporating the high-resolution acquired areas, displayed as rectangles. As all other elements in the reconstruction model, the areas can be loaded or revisited by clicking on them (Fig. 9).The total number of areas imaged was 160.

Analysis of Areas of Interest

HB. All 160 areas were segmented overnight using the Analyze Case function. In this particular case, DAPI counterstained nuclei were segmented using a two-step algorithm that first uses an adaptative threshold to separate DNA areas from the background and then applies a Hough-transform + Watershed algorithm to separate clusters of nuclei resulting of the adaptative thresholding. FISH signals were segmented using a TOP-HAT morphological algorithm followed by morphological reconstruction. The FISH segmentation algorithm was applied only to those areas identified as nuclei on the DAPI channel. A detailed description of these methods is out of the scope of this study and can be found in the literature (Malpica et al., 1997, Ortiz de Solorzano et al., 1998, Malpica et al., 1999).

The total number of nuclei segmented was above 200,000. The results of the nuclear and FISH segmentation can be displayed for every area (Fig. 9) from the images of the sections they belong to or directly from the 3D rendering window.

DISCUSSION

We have presented a new and powerful computer-based system that allows automation of the acquisition, storage, and analysis of thick sectioned tissue specimens. The system has been described and demonstrated by reconstructing tissue blocks from two tissue sources, processed using different staining and microscopy protocols.

By integrating information from different types of staining, both histological (e.g., H&E) and molecular (e.g., IHQ or FISH), we allow simultaneous morphological and molecular analysis of the specimens. The molecular analysis is not only simultaneous, but in fact driven by the morphology as the areas acquired and analyzed can be selected directly from the 3D recon-

struction of the structures marked on the low-resolution images. This way, the labor-intensive task of acquiring and analyzing the images is done semi-automatically—the slides still need to be manually placed on the microscope and the areas to be acquired drawn on the images of the sections—enormously reducing the time and labor. In fact, large-scale analysis can be done (e.g., whole gland analysis), which would be unthinkable otherwise.

A classical problem of fluorescence microscopy, which limits the number of labeled elements to the number of fluorochromes that can be discriminated from, can be overcome by using single or dual color staining on several adjacent sections, provided that the distance between histological (e.g., H&E) sections enclosing them allows detection of continuity between the structures of interest.

Future work will involve adding new functionality to the system by augmenting the number of analysis functions provided (e.g., detection of cytoplasmic or extra cellular proteins), and automating the detection of the structures of interest, which at this point is the most time-consuming task when reconstructing a Case. Interaction could be further reduced by using an automatic slide feeder attached to the microscope, which would eliminate the manual loading of slides for revisiting or acquisition of areas.

ACKNOWLEDGMENTS

We thank K. Chew and S. Ravani for their help selecting and coordinating tissue sectioning and staining, Drs. G. Maggrane and K. Ewan for their work on the IHQ and FISH staining of the sections, Drs. F. Waldman, K. Chin, and J. Gray for their support and advice regarding the selection of the tissue, and A. Harris for her help with tissue imaging and annotation.

REFERENCES

Boissonnat JD, Geiger B. 1993. Three dimensional reconstruction of complex shapes based on the *Delaunay* triangulation. In: Acharya RS, Goldgof DB, editors. Biomedical image processing and biomedical visualization. SPIE 1905:964–975.

Dean P, Mascio L, Ow D, Sudar D, Mullikin J. 1990. Proposed standard for image cytometry data files. Cytometry 11:561–569.

Firestone L, Cook K, Culp K, Talsania N, Preston K. 1991. Comparison of autofocus methods for automated microscopy. Cytometry 12:95–206.

Groen F, Young IT, Ligthhart G. 1985. A comparison of different focus functions for use in autofocus algorithms. Cytometry 12:81–91.

Malpica N, Ortiz de Solórzano C, Vaquero JJ, Santos A, Vallcorba I, García-Sagredo JM, del Pozo F. 1997. Applying watershed algorithms to the segmentation of clustered nuclei: defining strategies for nuclei and background marking. Cytometry 28:289–297.

Ortiz de Solórzano C, Santos A, Vallcorba I, García-Sagredo JM, del Pozo F. 1998. Automation of FISH Spot counting in interphase nuclei: statistical evaluation and data correction. Cytometry 31:93–99.

Ortiz de Solórzano C, García Rodriguez E, Jones A, Sudar D, Pinkel D, Gray JW, Lockett SJ. 1999. Segmentation of confocal microscope images of cell nuclei in thick tissue sections. J Microsc 193:212–226.

Santos A, Ortiz de Solorzano C, Vaquero JJ, Pena JM, Malpica N, del Pozo F. 1997. Evaluation of autofocus functions in molecular cytogenetic analysis. J Microsc 188:264–272.

Vollath D. 1987. Automatic focusing by correlative methods. J Microsc 147:249–288.

Journal of Biomedical Optics 9(3), 444–453 (May/June 2004)

Appendix 2

Automatic segmentation of histological structures in mammary gland tissue sections

R. Fernandez-Gonzalez
Lawrence Berkeley National Laboratory
Life Sciences Division
1 Cyclotron Road, Building 84, MS 84-171
Berkeley, California 94720
and
University of California,
Berkeley & University of California,
San Francisco Joint
Graduate Group in Bioengineering
467 Evans Hall #1762
Berkeley, California 94720-1762
E-mail: RFGonzalez@lbl.gov

T. Deschamps
University of California, Berkeley
Mathematics Department
970 Evans Hall #3840
Berkeley, California 94720-3840
and
Lawrence Berkeley National Laboratory
Mathematics Department
Berkeley, California 94720

A. Idica
Lawrence Berkeley National Laboratory
Life Sciences Division
1 Cyclotron Road, Building 84, MS 84-171
Berkeley, California 94720

R. Malladi
Lawrence Berkeley National Laboratory
Mathematics Department
Berkeley, California 94720

C. Ortiz de Solorzano
Lawrence Berkeley National Laboratory
Life Sciences Division
1 Cyclotron Road, Building 84, MS 84-171
Berkeley, California 94720

Abstract. Real-time three-dimensional (3-D) reconstruction of epithelial structures in human mammary gland tissue blocks mapped with selected markers would be an extremely helpful tool for diagnosing breast cancer and planning treatment. Besides its clear clinical application, this tool could also shed a great deal of light on the molecular basis of the initiation and progression of breast cancer. We present a framework for real-time segmentation of epithelial structures in two-dimensional (2-D) images of sections of normal and neoplastic mammary gland tissue blocks. Complete 3-D rendering of the tissue can then be done by surface rendering of the structures detected in consecutive sections of the blocks. Paraffin-embedded or frozen tissue blocks are first sliced and sections are stained with hematoxylin and eosin. The sections are then imaged using conventional bright-field microscopy and their background corrected using a phantom image. We then use the fast-marching algorithm to roughly extract the contours of the different morphological structures in the images. The result is then refined with the level-set method, which converges to an accurate (subpixel) solution for the segmentation problem. Finally, our system stacks together the 2-D results obtained in order to reconstruct a 3-D representation of the entire tissue block under study. Our method is illustrated with results from the segmentation of human and mouse mammary gland tissue samples. [DOI: 10.1117/1.1699011]

Keywords: breast cancer, molecular analysis, automatic segmentation, 3-D reconstruction, fast-marching, level set.

Paper 44006 received Jul. 1, 2003; revised manuscript received Sep. 8, 2003; accepted for publication Sep. 30, 2003. This paper is a revision of a paper presented at the SPIE conference on Three-Dimensional and Multidimensional Microscopy: Image Acquisition and Processing X, January 2003, San Jose, Calif. The paper presented there appears (unrefereed) in SPIE Proceedings Vol. 4964.

1 Introduction

The mature mammary gland is a treelike organ made up of three different levels of branching ducts converging at the nipple.[1] The ducts are lined by epithelial cells and end in secretory lobuloalveolar structures, which are the sites of milk production during lactation. In cancer, this hierarchical organization is disrupted by uncontrolled growth of the epithelium, and sometimes by the subsequent invasion of the surrounding tissue.[2] These morphological or structural changes are accompanied by other genetic and epigenetic changes at the cellular level (see Fig. 1). An example of this is seen in ductal carcinoma *in situ* (DCIS), which is a preinvasive form of cancer. In DCIS, the molecular changes common in breast cancer can be observed together with well-defined morphological patterns, such as comedo or cribiform ones, with or without central necrosis.[3,4]

An interesting problem is the quantification of these genetic or molecular changes in the context of the tissue environment where they occur. In this way, by looking at cancer as an organ that is inherently heterogeneous and dynamic, we believe that we will obtain a better understanding of the events that drive the progression of the disease. Therefore, since the normal mammary gland and its neoplastic variants are neither flat nor homogeneous, the approach to this problem should be three-dimensional as well, and take into account the heterogeneity of the gland. However, most classic methods in biology focus on only a single type of abnormality (molecular or morphological) and/or neglect three-dimensionality and heterogeneity.

Although imaging of both tissue structure and function *in vivo* would be extremely desirable, the existing *in vivo* imag-

Address all correspondence to R. Fernandez-Gonzalez. Lawrence Berkeley National Laboratory, Life Sciences Division, 1 Cyclotron Road, Building 84, MS 84-171, Berkeley, California 94720. Tel: 510-486-5359; Email: RFGonzalez@lbl.gov

1083-3668/2004/$15.00

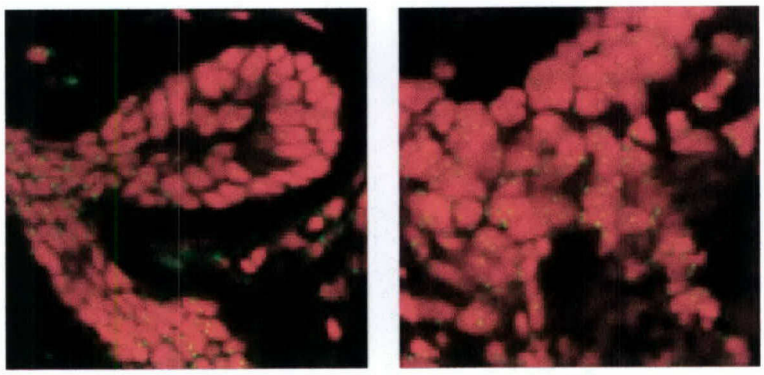


Fig. 1 Morphological and genetic alterations in breast cancer. The images show two optical sections of human mammary gland tissue acquired using a confocal laser scanning microscope; nuclei are displayed in red; a probe for a certain DNA sequence in chromosome 17 is shown in green. (a) Normal tissue; as expected, each nucleus contains up to two green signals (up to two copies of chromosome 17 per nucleus in a single optical section). (b) Neoplastic lesion; not only do some nuclei contain more than two copies of the probe, but they also have distinct morphological changes that can be observed in this section.

ing methods (x-ray, magnetic resonance imaging, optical tomography, etc.) do not provide the necessary resolution for cell-level molecular analysis. In addition, these methods provide morphological information that can only be indirectly related to the function of the tissue. Consequently, *ex vivo* microscopic analysis of the tissue from flat fixed sections is the routine method in histopathology. However, the limited ability of the human eye to extrapolate and visualize 3-D structures from sequences of 2-D scenes renders this method unsuitable for quantitative 3-D tissue characterization. To overcome these issues, we have developed a system for simultaneous morphological and molecular analysis of thick tissue samples.[5] The system consists of a computer-assisted microscope and a JAVA-based image display, analysis, and visualization program (R3D2). R3D2 allows semiautomatic acquisition, annotation, storage, 3-D reconstruction, and analysis of histological structures (intraductal tumors, normal ducts, blood vessels, etc.) from thick tissue specimens. For this purpose the tissue needs to be embedded in a permanent or semirigid medium after collection. The tissue is then fully sectioned, and the resulting sections are stained in a way that visually highlights the desired structures. In histopathology, hematoxylin and eosin (H&E) is the most common combination of dyes used to observe the morphology of the tissue. In order to image not only structure, but also molecular events, we alternate H&E staining with fluorescent staining (immunofluorescence, fluorescence *in situ* hybridization) of proteins and selected genes in consecutive sections.

This paper focuses on the annotation of the structures of interest on the H&E-stained sections. Manual segmentation has been used before to delineate histological structures.[6–8] In order to build the 3-D reconstruction of the block, we initially annotated each of the interesting features of the tissue on each section by using manually drawn contours. This step constituted a bottleneck in the study of samples. Semiautomatic approaches to segmentation of features of interest in histological sections have also been used,[9] but they still involve too much user interaction to be useful for reconstructing large tissue samples. Automatic segmentation of the structures of interest is the answer to this problem. We propose an automatic method followed by interactive correction that greatly reduces the amount of interaction required, thus allowing us to use our system for imaging and reconstruction of large, complex tissue structures.

Automatic extraction of contours in 2-D images is usually done with active contour models, originally presented by Kass et al.[10] These methods are based on deforming an initial contour (polygon) toward the boundary of the desired object to extract in an image. The deformation is achieved by minimizing a certain energy function, which is computed by integrating along the contour, terms related to its continuity, and terms related to the pixel values of the area of the image where the contour is defined. That energy function approaches a minimum near the object's boundary, and thus the minimization process drives the curve toward the desired shape.

As an alternative, implicit surface evolution models have been introduced by Malladi et al.[11,12] and Caselles et al.[13] In these models, the curve and surface models evolve under an implicit speed law containing two terms, one that attracts it to the object's boundary and another that is closely related to the regularity of the shape. Specifically, the proposal is to use the level-set approach of Osher and Sethian.[14] This is an interface propagation technique used for a variety of applications, including segmentation. The initial curve is represented here as the zero level set of a higher dimensional function, and the motion of the curve is embedded within the motion of that higher dimensional function. An energy formulation similar to the active contours leads to a minimization process with several advantages. First, the zero level set of the higher dimensional function is allowed to change topology and form sharp corners. Second, geometric quantities such as normal and curvature are easy to extract from the hypersurface. Finally, the method expands straightforwardly to 3-D,[15] but adding a dimension to the problem increases the computational cost associated with the method. The narrow-band approach of Adalsteinsson and Sethian[16] accelerates the level-set flow by considering for computations only a narrow band of pixels around the zero level set. However, in our experience, the

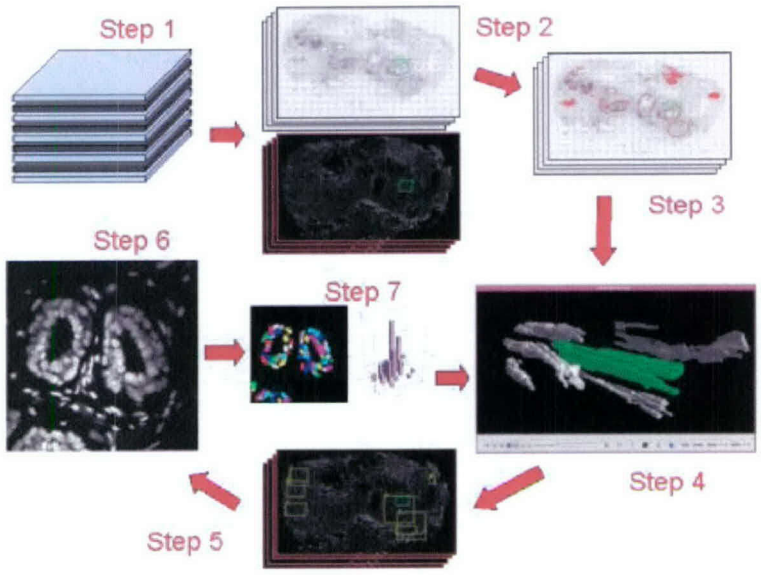


Fig. 2 Protocol followed on tissue blocks. The different steps (sectioning, annotation, reconstruction, high-magnification acquisition, and molecular analysis) are illustrated. Samples are fully sectioned at 5 μm (step 1). The odd sections are stained with H&E, the even ones with some kind of fluorescence technique (application dependent). Images are acquired of all the sections (step 2), and the structures of interest are delineated in the H&E-stained ones (step 3). A 3-D reconstruction of the specimen is created from these markings (step 4). From the 3-D reconstruction of the tissue, different areas can be selected for molecular analysis (step 5). The system will take high-magnification images of those areas on the corresponding fluorescent sections (step 6). Image analysis tools can then be used to quantify the presence and distribution of molecular markers in the high-magnification images (step 7).

narrow-band technique does not reduce the computational cost to a reasonable limit, owing to the large size of the images of the sections.

Thus we propose to use the fast-marching method.[17] This method considers monotonically advancing fronts (speed always positive or negative), providing a result very quickly, albeit one that is not as accurate as the one obtained by using the level-set algorithm. Malladi and Sethian[18] showed that the fast-marching method can be used as the initial condition for the slower but more accurate level-set segmentation, obtaining real-time delineation of the structures of interest. A combination of all these tools is the framework we use to reconstruct normal and cancerous ducts in mammary gland tissue sections of DCIS samples.

This paper is organized as follows. Section 2 describes the general tissue handling and image acquisition protocols that we use, as well as the theoretical basis of the segmentation scheme; Sec. 3 shows the results of applying our method to histological tissue sections; and Sec. 4 discusses the results and suggests some future developments and improvements to our approach.

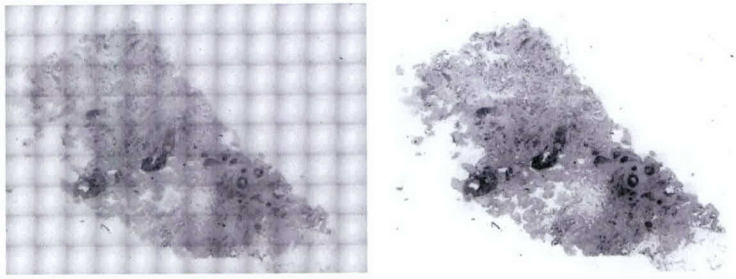
2 Methodology

2.1 *Tissue Processing and Imaging*

The tissue processing and staining protocol used is illustrated in Fig. 2. Tissue blocks of 4- to 5-mm thickness were sliced into 5 μm (thin) sections (step 1). The odd sections were stained with H&E to obtain morphological information at both the cytological (single cell) and architectural (organ) levels. The even sections were stained using some fluorescence technique (e.g., immunocytochemistry, fluorescence *in situ* hybridization), depending on the molecular phenomena that we wanted to study. Describing the acquisition and analysis of the fluorescent images is out of the scope of this paper, which focuses on the segmentation of epithelial structures in the H&E-stained sections. Therefore the rest of this section describes only the protocol used for the H&E-stained sections. Low magnification (2.5×), panoramic images of all the sections were automatically acquired using a motorized Zeiss Axioplan I microscope coupled with a monochrome XilliX Microimager CCD camera (step 2). To create these large panoramic images, the system scanned the entire section, taking single-field-of-view snapshots and tiling them together into single, whole-view images of the sections. The required sequence of microscope movements and camera operations is produced by an application running on the Sun Ultra 10 workstation that controls the camera and all moving parts of the microscope.

Next we annotated the structures of interest (ducts, lymph nodes, tumors) in the images of the H&E-stained sections (step 3). These annotated structures were used to produce a three-dimensional model of each tissue block (step 4).

These four initial steps are the focus of this paper. However, to better understand the rationale for the process, we briefly describe the final three steps, which are out of the intended scope for this paper. The user can choose to revisit areas in the three-dimensional rendition of the organ, based on their morphology. This can be done on the H&E sections or on the intermediate fluorescent sections. To do so, the system requires the user to acquire high-magnification (40 to 100×) images of the corresponding section(s) (steps 5 and 6). If the high-magnification images are taken from the intermediate



(a) (b)

Fig. 3 Background correction. (a) Image of a section belonging to a human case. The background pattern created by the acquisition method is readily noticeable. (b) Same image after background correction.

immunofluorescent sections, quantification routines can be run on these new images (step 7).

Manually annotating all the relevant morphological structures in all the sections of fully sectioned tissue blocks is feasible but, for all purposes, impractical because of the tremendous human effort needed. Although it might be the most accurate and reliable approach, manual annotation is not possible except when reconstructing small, very simple tissue volumes. As a result, we have developed automatic methods that eliminate or greatly reduce human interaction, thus making the reconstruction of complex systems possible. The following discussion describes our approach.

2.2 *Segmentation*

2.2.1 *Background removal*

The algorithm that we use to acquire an image of an entire section creates a mosaic from a set of snapshots (one per field of view). This approach gives rise to a background pattern across the image [Fig. 3(a)] involving relatively large gradients in between elements of the mosaic. The objects of interest often span several fields of view, and since our segmentation approach depends largely on the gradients of the image, we need to eliminate the background pattern in order to obtain good segmentation results. This can be done by performing a set of arithmetic operations on the "mosaic" image, known as background compensation.

First we need a phantom, that is, an image of an empty field of view taken under the same illumination conditions and microscope configuration that we used to acquire the initial image. Since most of the images that we acquire have an empty frame in the upper left corner, it is simple to choose that frame as our phantom for the corresponding section. After normalizing the pixel values in the phantom, and for each frame in the entire image, we divide the value at each pixel by the value at the corresponding pixel in the phantom frame. The resulting image is background-corrected as shown in Fig. 3(b), and it is a better input for our segmentation algorithms.

2.2.2 *Preliminary segmentation: fast-marching method*

Consider a monotonically advancing 2-D front C with a speed F that is always positive in the normal direction, starting from an initial point p_0,

$$\frac{\partial C}{\partial t} = F\mathbf{n}. \tag{1}$$

This equation drives the evolution of a front starting from an infinitesimal circular shape around p_0 until each point p inside the image domain is visited and assigned a crossing time $U(p)$, which is the time t at which the front reaches point p.

The gradient of the arrival time is inversely proportional to the speed function, and thus we have a form of the eikonal equation

$$|\nabla U|F = 1 \quad \text{and} \quad U(p_0) = 0. \tag{2}$$

One way to solve Eq. (2) is to use upwind finite-difference schemes and iterate the solution in time.[15] In other words, the scheme relies on one-sided differences that look in the upwind direction of the moving front, thereby choosing the correct viscosity solution, namely

$$[\max(u - U_{i-1,j}, u - U_{i+1,j}, 0)]^2 + [\max(u - U_{i,j-1}, u - U_{i,j+1}, 0)]^2 = \frac{1}{F_{i,j}^2}. \tag{3}$$

The key to solving this equation rapidly is to observe that the information propagates from smaller values of U to larger values in the upwind difference structure in Eq. (3). The idea is to construct the time surface, one piece at a time, by only considering the "frontier" points; we detail the fast-marching method in Table 1.

Note that in solving Eq. (3), only alive points are considered. This means that for each point, the calculation is made using the current values of U at the neighbors, and not estimates at other trial points. Considering the neighbors of the grid point (i,j) in 4-connectedness, we designate $\{A_1, A_2\}$ and $\{B_1, B_2\}$ as the two couples of opposite neighbors so that we get the ordering $U(A_1) \leq U(A_2)$, $U(B_1) \leq U(B_2)$, and $U(A_1) \leq U(B_1)$. Since we have $u \geq U(B_1) \geq U(A_1)$, we can derive

$$[u - U(A_1)]^2 + [u - U(B_1)]^2 = \frac{1}{F_{i,j}^2}. \tag{4}$$

Computing the discriminant (Δ) of Eq. (4), we complete the steps described in Table 2. Thus the algorithm needs only one

Table 1 Fast-marching algorithm.

Algorithm for 2-D Fast Marching

- Definitions:

  Alive set: all grid points where the action value U has been reached and will not be changed;

  Trial set: next grid points (4-connectedness neighbors) to be examined. An estimate U of U has been computed using Eq. (3) from Alive points only (i.e., from U);

  Far set: all other grid points, where there is no estimate for U yet;

- Initialization:

  Alive set: reduced to the starting point p_0, with $U(p_0) = U(p_0) = 0$;

  Trial set: reduced to the four neighbors p of p_0 with initial value $U(p) = 1/F(p)$ ($U(p) = \infty$);

  Far set: all other grid points, with $U = \infty$;

- Loop:

  Let $p = (i_{\min}, j_{\min})$ be the trial point with the smallest action U;

  Move it from the trial to the alive set (i.e., $U(p) = U_{i_{\min}, j_{\min}}$ is frozen);

  For each neighbor (i, j) (4-connectedness in 2-D) of $(i_{\min}, j_{\min})$;

  If (i, j) is far, add it to the trial set and compute $U_{i,j}$ using Eq. (3);

  If (i, j) is trial, update the action $U_{i,j}$ using Eq. (3).

pass over the image to find a solution. To execute all the operations that we just described in the minimum amount of time, the trial points are stored in a min-heap data structure.[17] Since the complexity of changing the value of one element of the heap is bounded by a worst-case processing time of $[O(\log_2 N)]$, the total algorithm has a complexity of $O(N \log_2 N)$ on a grid with N nodes.

Table 2 Solving the upwind scheme locally.

1. ● If $\Delta \geq 0$, u should be the largest solution of Eq. (4);

   If the hypothesis $u > U(B_1)$ is wrong, go to 2;

   If this value is larger than $U(B_1)$, this is the solution;

   ● If $\Delta < 0$, B_1 has an action too large to influence the solution. It means that $u > U(B_1)$ is false. Go to 2;

   Simple calculus can replace case 1 by the test:

   If $1/F_{i,j} > U(B_1) - U(A_1)$, $u = U(B_1) + U(A_1) + \{2 1/F_{i,j}^2 - [U(B_1) - U(A_1)]^2\}^{1/2}/2$ is the largest solution of Eq. (4) else go to 2;

2. Considering that we have $u < U(B_1)$ and $u \geq U(A_1)$, we finally have $u = U(A_1) + 1/F_{i,j}$.

Finally, we define the speed of propagation in the normal direction as a decreasing function of the gradient $|\nabla I(x)|$, that is, a function that is very small fnear large image gradients (i.e., possible edges) and large when the brightness level is constant:

$$F(x) = \exp - \alpha |\nabla I(x)|, \quad \alpha > 0, \tag{5}$$

where α is the edge strength, or the weight that we give to the presence of a gradient in order to slow down the front. Depending on this value, the speed function falls to zero more or less rapidly, and thus it could stop a few grid points away from the real edge. Also, variations in the gradient along the boundary can cause inaccurate results. False gradients caused by noise can be avoided using an edge-preserving smoothing scheme on the image as a preprocessing step; see Ref. 19.

The user can run the fast-marching method from a given set of initial points (mouse clicks on the background of the images). Alternatively, the user can decide to segment only the structures within a manually defined rectangular region of interest. If the region is too big, subsampling can be used so that the segmentation process is not too slow. However, this option must be used carefully, since subsampling smooths the boundaries of the objects present in the image and can completely obliterate smaller structures. The resulting contour (or contours if we are segmenting several objects at the same time) will provide an excellent initial condition for the level-set method.

Putting all of these elements together, we are able to obtain a good approximation of the shape of the object that we are trying to segment (Fig. 5). In order to improve the final result, we propose to run a few iterations of the level-set method using the result of the fast-marching method as the initial condition.

2.2.3 *Final segmentation: level-set method*

Once we have obtained a good approximation of the shape of the object using the fast-marching algorithm, we can afford to use the more computationally expensive level-set method to improve the result of the segmentation. The essential idea here is to embed our marching front as the zero level set of a higher dimensional function. In our case, we take that function to be $\phi(x) = \pm d$, where d is the signed distance from x to the front (see Fig. 4), assigning negative distances for pixels inside the evolving curve, and positive distances for pixels outside it.

Following the arguments in Ref. 14, we obtain the following evolution equation:

$$\phi_t + F(x,y)|\nabla \phi| = 0, \tag{6}$$

where F is again the speed in the normal direction. This is an initial-value partial differential equation, since it describes the evolution of the solution on the basis of an initial condition defined as $\phi(x, t = 0) = \phi_0$. As pointed out earlier, the level-set approach offers several advantages:

- The zero level set of the function can change topology and form sharp corners.

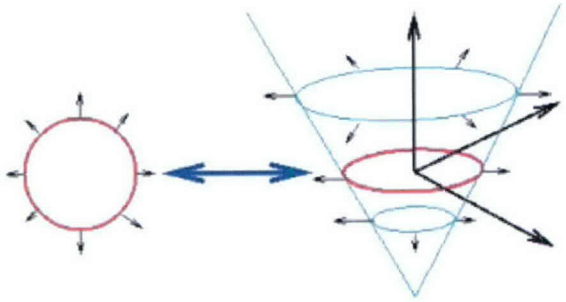


Fig. 4 Basic concept behind the level-set method. The marching front is embedded as the zero level set of a higher dimensional function.

- A discrete grid can be used together with finite differences to approximate the solution.
- Intrinsic geometric quantities such as normal and curvature can be easily extracted from the higher dimensional function.
- Everything extends directly to 3-D.

To mold the initial condition (in our case the result of the fast-marching method) into the desired shape, we use two force terms. By substituting these two terms in the motion equation[18,20] we get:

$$\phi_t - g(1-\epsilon\kappa)|\nabla\phi| - \beta\nabla g \times \nabla\phi = 0, \quad \epsilon > 0, \quad \beta > 0, \tag{7}$$

where g is an edge indicator function defined by the speed of the front [Eq. (5)] in the eikonal Eq. (2); κ is the curvature of the expanding front, and ϕ_t is the unknown that we are trying to compute. As before, the image $I(x)$ can be preprocessed using an edge-preserving smoothing scheme.

The second term of the equation has two components. The first one slows the surface in the neighborhood of high gradients (edges), while the second one (motion by curvature) provides regularity to the curve. The parameter ϵ is the weight of the motion by curvature term, and determines the strength of its regulatory effect: the bigger we make ϵ, the more we limit the possibility of obtaining sharp corners and irregular contours. In practice, an intermediate value of ϵ provides a good trade off between contour smoothing and accuracy.

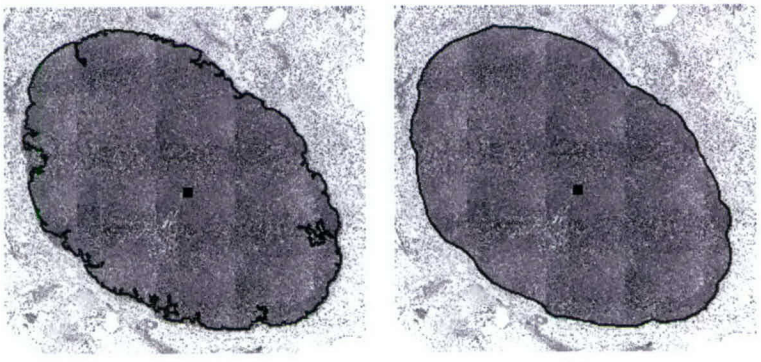
Finally, the last term of the equation attracts the front to the object's boundaries. It aligns all the level sets with the ideal gradient, which would be a step function centered at the point of maximum gradient in the original image. β is the weight of the advection of the front by the edge vector field ∇g. It determines the strength of the attraction of the front to the edges.

At times, for very small objects, it is possible to use the level-set method from the initial point. However, for any type of morphological structure, we can use the result of the eikonal Eq. (2), $U(x,y)$, as the initial condition for the level-set algorithm: $\phi(x,y;t=0)=U(x,y)$. Then by solving Eq. (7) for a few time steps using the narrow-band approach, we obtain an accurate, real-time segmentation of the desired object (see Fig. 5).

3 Results

In this section we consider the problem of reconstructing DCIS areas in a tissue biopsy of a cancerous human breast, as well as a group of normal ducts through an entire mouse mammary gland. The tissue samples were sliced for a total of 55 sections in the human tissue and 206 sections in the mouse one. We used all of the sections for the reconstruction of the human case and 40 sections for reconstruction of the mouse case. Manually delineating all the structures in every section is an extremely time-consuming process. To automatically segment those structures using the framework described in Sec. 2, we begin by defining a region-of-interest (ROI) where we will run the segmentation algorithm. This ROI can be extended to cover the entire section, but considering the size of the images, it is wise to use subsampling in order to run the algorithm in real time. The level of subsampling can be determined by the user; greater subsampling can be used on large ROIs without compromising the resolution and accuracy of the final segmentation.

After defining the ROI, we have to tune the different parameters of the segmentation process, particularly α [Eq. (5)]



(a) (b)

Fig. 5 The segmentation of a lymph node in a mouse mammary gland section is shown in black. (a) The result of the fast-marching method; it provides a good approximation to the boundaries of the lymph node; the blue point in the middle of the lymph node is the initial contour from which the algorithm was run. (b) The result of using the level-set method after the fast-marching method; the final contour is more accurate and smoother. In both cases the images were subsampled in the x and y directions to be able to run the segmentation in real time.

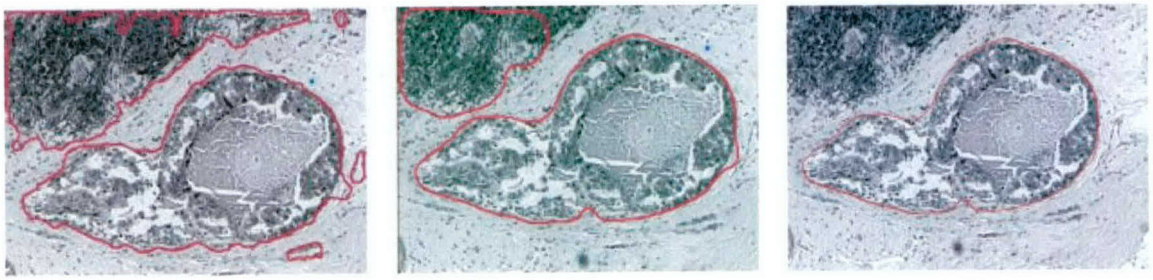


Fig. 6 Segmentation of a DCIS tumor in human mammary gland tissue. (a) Segmentation using just the level set method. The blue dot represents the initial contour. (b) Results of the full segmentation with blue lines delineating tumor masses. (c) Results after editing using the interactive tools provided by the system.

and ϵ [Eq. (7)]. This is done by the user based on the default values provided by the algorithm and the type of object that he or she is trying to segment. However, we have observed that a particular set of parameters is frequently good enough to segment similar structures (i.e., all the tumors, all the ducts, all the lymph nodes) throughout all the sections of a particular tissue block. Thus the user only needs to modify the parameters the first time that he or she tries to segment a new type of morphological element in the tissue.

After selecting the parameters of the flow, initial points are defined inside (to find the internal contour) or outside (to find the external contour) the structures of interest. In most cases one computer mouse click is enough, although large images may require several evenly distributed clicks. The value of $U(x)$ at these points is set to zero as in Eq. (2), and the fast-marching algorithm of Table 1 is executed. When this method finishes, the final $U(x)$ function is passed as the initial condition to the level-set motion Eq. (7). We then iterate this equation for a few steps. This segmentation scheme provides a result in less than 1 s for images whose size (after subsampling, if any) is around 2 kbytes (e.g., 512×512 pixels) running on a Sun Ultra 10 workstation with 1 Gbytes of RAM.

3.1 *Human Case Segmentation*

Figure 6 displays the result obtained with the combination of the fast-marching and the level-set methods in a tissue biopsy from a patient with an intraductal carcinoma. Figure 6(b) shows the segmentation of a tumor mass in a human tissue block. The results of the segmentation can be edited and removed with the interactive tools provided by our system [see Fig. 6(c)]. For this segmentation, an area was selected around the structures of interest and no subsampling was used. The initial contours are represented by blue points in Fig. 6(a).

3.2 *Mouse Case Segmentation*

In Fig. 7 we can see the segmentation of the external contours of several ducts in a particular area of a mouse mammary gland. In this case we also run the algorithm on an ROI on one of the sections with no subsampling factor. Figure 7(a) displays the points where we initialized the contours. Figures 7(b) and 7(c) show the results of the segmentation before and after interactive correction of the results, respectively.

3.3 *3-D Reconstruction*

Finally the segmented shapes are connected (manually) between sections, and 3-D reconstructions of the samples are built. Figure 8 shows a reconstruction of the tumors contained in the human tissue block, including the tumor shown in Fig. 6. Increasing the "motion by curvature" term, as described in Sec. 2, can reduce surface noise. In Fig. 9 we can see the reconstruction of the normal ducts segmented in the images of the mouse mammary gland (see one of the corresponding 2-D segmentations in Fig. 7).

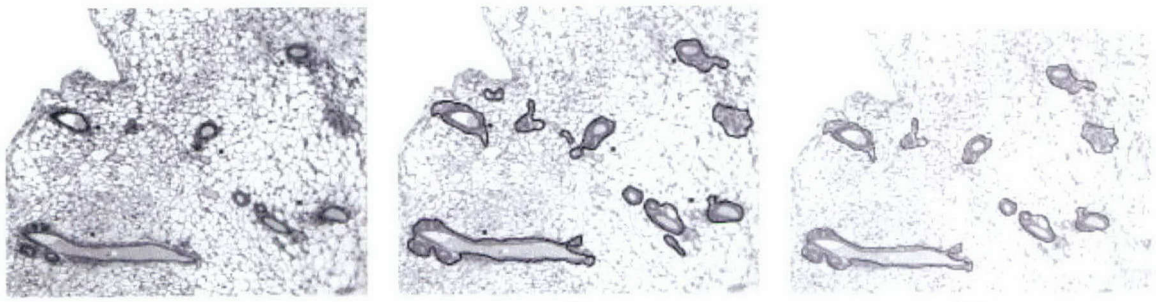


Fig. 7 Segmentation of normal ducts in a mouse mammary gland tissue sample. (a) Initial data. (b) Results of the segmentation (red lines delineate normal ducts). (c) Results after editing using the interactive tools provided by the system.

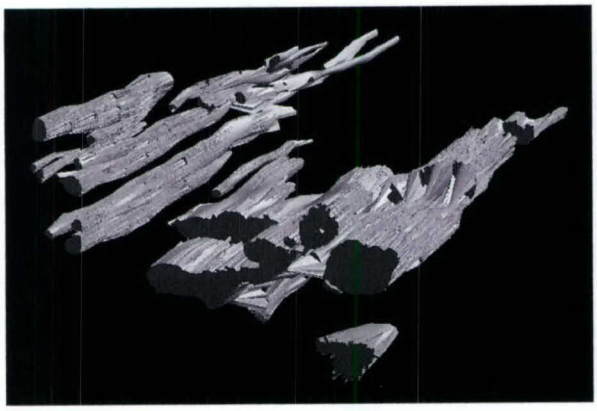


Fig. 8 3-D reconstruction of tumors in a human mammary gland tissue block. Tumor masses are rendered as gray volumes. The scene was stretched ten times in the z direction to obtain a better view.

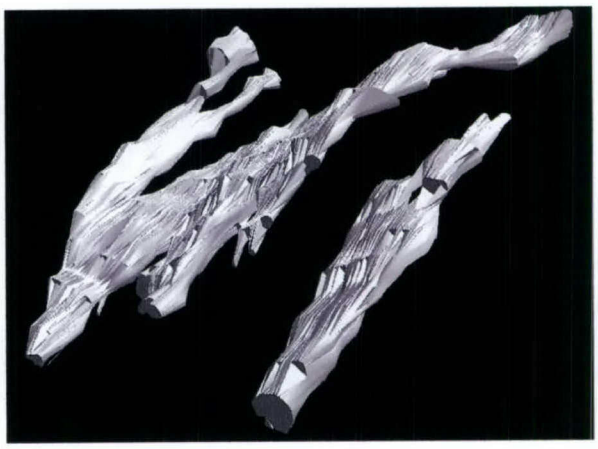


Fig. 9 3-D reconstruction of normal ducts in a mouse mammary gland. The ducts are rendered as gray volumes. A single duct and its branches can be traced throughout the gland. The z direction was not stretched in this case.

3.4 *Further Examples*

Figures 10 and 11 show two more examples of the results that can be obtained with our segmentation approach. In both cases the initial ROI was subsampled by a factor of 2 in both the x and y directions. The same segmentation parameters (α and ϵ) were used for both examples.

Figure 10(a) shows a DCIS lesion together with a normal duct in a human tissue section. Both structures were segmented at the same time using a single initial point, as can be seen in Fig. 10(b). Figure 10(c) shows the results incorporated to the full resolution image.

In Fig. 11(a), a terminal ductal lobular unit (TDLU) can be observed. These are lobuloalveolar structures where milk is produced during lactation in the human breast. The multiple alveoli that form the TDLU, together with the presence of a ductal part, make automatic segmentation of this type of structure a difficult task. However, after subsampling the image, our algorithm is able to find a contour that surrounds the entire structure [Figs. 11(b) and 11(c)], thus allowing its 3-D reconstruction.

4 Discussion

We have developed a microscopy system that semiautomatically reconstructs histological structures from fully sectioned tissue samples. However, the interaction required to operate the system is quite intensive, limiting the scope of its application to small tissue volumes or to studies not requiring high throughput analysis of the samples. In this paper we have presented a method that reduces the time and interaction needed to build the 3-D reconstruction of a tissue block, thus increasing the potential throughput of the system and therefore allowing us to use this approach for the reconstruction of large, complex specimens. To achieve this goal we have combined image processing techniques and two well-established schemes for interface propagation: the fast-marching method and the level-set method.

Our approach starts by correcting the background of the images. This is an important step, since the background pattern generated during image acquisition modifies the gradient

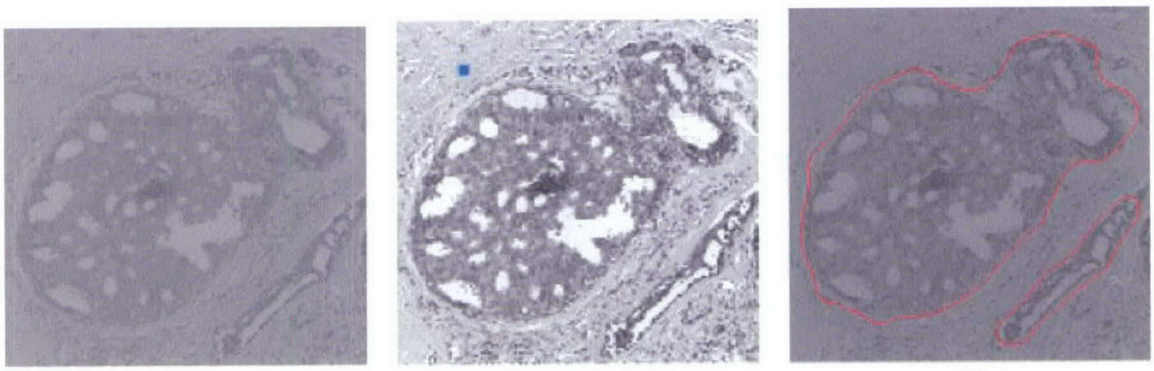


Fig. 10 Segmentation of a different DCIS tumor in human mammary gland tissue. (a) The duct on the-left contains a DCIS lesion with a necrotic center; the one on the right is normal. (b) The ROI was subsampled by a factor of 2 in both the x and y directions; the blue dot represents the initial seed. (c) Results of the segmentation on the full-resolution image.

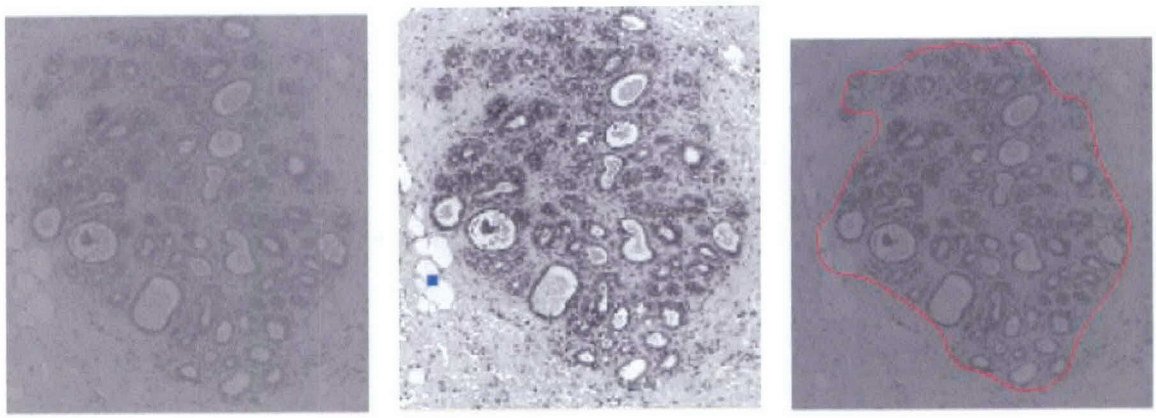


Fig. 11 Segmentation of a terminal ductal lobular unit in a human mammary gland. (a) The duct on the left contains a section of a TDLU, one of the sites of milk production in the human breast. (b) The ROI was subsampled by a factor of 2 in both the x and y directions; the blue dot represents the initial seed. (c) Results of the segmentation on the full-resolution image.

of the image, and the speed function that we use for interface propagation depends on that gradient. Once the background has been corrected, we run the fast-marching method. This technique provides a good approximation of the boundaries of the objects that we are trying to segment in a very short time, since it assumes monotonic speed functions (always positive or always negative). We then use the approximation provided by the fast-marching method as the initial condition for the level-set method. This more computationally expensive algorithm is run for just a few steps, enough to fit the front to the contours of the structures of interest, but not as many as to make the segmentation too time-consuming.

Though this approach is very useful, it can still be improved. The most accurate segmentations (and reconstructions) are obtained on full-resolution images. However, for some large structures, such as lymph nodes, or when trying to delineate multiple elements at the same time (for example, a group of ducts), segmentation on a full-resolution image is not real time, and can take up to 1 min. Using subsampling takes the segmentation execution time back to real time at the expense of some accuracy loss. Also, in areas with a lot of texture in the tissue around the ducts (stroma), tuning the parameters of the algorithm (α and ϵ) becomes more difficult. At times this process can take a few trials, since the expanding front tends to get trapped in high gradients that do not correspond to the boundaries of the feature that we are trying to segment, but to the texture of the stroma. Finally, once all the structures of interest have been segmented, the user still needs to manually connect them between sections. This constitutes a new bottleneck in the tissue analysis process.

For these reasons, we are currently working on a time step-independent scheme that is expected to be faster than the current one. To improve the accuracy of the results, we have developed an edge-preserving smoothing algorithm based on the Beltrami flow,[19] which can replace the Gaussian smoothing currently used before executing the segmentation methods. This algorithm eliminates false gradients that are due to noise, while enhancing gradients that are due to the object's boundaries, thus allowing the front to fit the boundaries of the object more accurately. Finally, the level-set approach is readily extensible to 3-D. The ability to segment 3-D structures of interest versus 2-D ones would save the process of connecting the segmented 2-D contours from section to section, thus improving the analysis time. Also, since geometric properties can be easily extracted from the higher dimensional function used in the level-set algorithm, we could readily obtain some information about the extracted volume from the segmentation algorithm itself. We are also working on a fully interactive segmentation method based on the model of the intelligent scissors,[21] because the selection of the seed points for segmentation is something that cannot be automated, owing to the high variability in the shapes of object's to be segmented.

In conclusion, we have presented a real-time method for automatic segmentation of morphological structures in mammary gland tissue sections. It is precisely the delineation of those structures that required the heaviest user interaction in our sample analysis protocol. Therefore, the automatic approach to segmentation that we describe here represents a first step toward real-time reconstruction and analysis of mammary gland samples. Achieving that goal would allow us to accelerate our studies on the biological basis of human breast cancer. Moreover, obtaining real-time reconstruction and analysis of samples from our system would be useful for pathological diagnosis in a clinical environment; 3-D renderings of all the morphological structures in a mammary gland biopsy could be mapped with the distribution of particular markers of breast cancer within a few hours of extracting the tissue from the patient. From an intrasurgical point of view, the renderings would prove—tissue processing and stain permitting—an important tool in the evaluation of breast tumors and their margins.

Acknowledgments

The U.S. Army Medical Research Materiel Command under grants DAMD17-00-1-0306 and DAMD17-00-1-0227, the Lawrence Berkeley National Laboratory Directed Research

and Development Program, and the California Breast Cancer Research Program, through project #8WB-0150, supported this work. This work was also supported by the Director, Office of Science, Office of Advanced Scientific Research, Mathematical, Information, and Computational Sciences Division, U.S. Department of Energy under contract no. DE-AC03-76SF00098.

References

1. L. Hennighausen and G. W. Robinson, "Think globally, act locally: the making of a mouse mammary gland," *Genes Dev.* **12**(4), 449–455 (1998).
2. L. M. Franks and N. M. Teich, *Introduction to the Cellular and Molecular Biology of Cancer*, Oxford University Press, Oxford (1997).
3. T. Tot, L. Tabar, and P. B. Dean, *Practical Breast Pathology*, 1st ed. Falun Central Hospital, Thieme (2002).
4. R. D. Cardiff and S. R. Wellings, "The comparative pathology of human and mouse mammary glands," *J. Mammary Gland Biol. Neoplasia* **4**(1), 105–122 (1999).
5. R. Fernandez-Gonzalez, A. Jones, E. Garcia-Rodriguez, P. Y. Chen, A. Idica, M. H. Barcellos-Hoff, and C. Ortiz de Solorzano, "A system for combined three-dimensional morphological and molecular analysis of thick tissue samples," *Microsc. Res. Tech.* **59**(6), 522–530 (2002).
6. T. Ohtake, R. Abe, I. Kimijima, T. Kukushima, A. Tsuchiya, K. Hoshi, and H. Wakasa, "Computer graphic three-dimensional reconstruction of the mammary duct-lobular systems," *Cancer* **76**(1), 32–45 (1995).
7. D. F. Moffat and J. J. Going, "Three-dimensional anatomy of complete duct systems in human breast: pathological and developmental implications," *J. Clin. Path.* **49**, 48–52 (1996).
8. T. Ohtake, I. Kimijima, T. Fukushima, M. Yasuda, K. Sekikawa, S. Takenoshita, and R. Abe, "Computer-assisted complete three-dimensional reconstruction of the mammary gland ductal/lobular systems," *Cancer* **91**(12), (2001).
9. F. Manconi, R. Markham, G. Cox, E. Kable, and I. S. Fraser, "Computer-generated, three-dimensional reconstruction of histological parallel serial sections displaying microvascular and glandular structures in human endometrium," *Micron* **32**, 449–453 (2001).
10. M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: active contour models," *Int. J. Comp. Vis.* **1**(4), 321–331 (1988).
11. R. Malladi, J. A. Sethian, and B. C. Vemuri, "A topology-independent shape modeling scheme," *Proc. SPIE* **2031**, 246–258 (1993).
12. R. Malladi, J. A. Sethian, and B. C. Vemuri, "Shape modelling with front propagation: a level set approach," *IEEE Trans. Pattern Anal. Mach. Intell.* **17**(2), 158–175 (1995).
13. V. Caselles, F. Catté, T. Coll, and F. Dibos, "A geometric model for active contours," *Numer. Math.* **66**, 1–31 (1993).
14. S. Osher and J. A. Sethian, "Fronts propagating with curvature-dependent speed: algorithms based on the Hamilton-Jacobi formulation," *J. Comput. Phys.* **79**, 12–49 (1988).
15. J. A. Sethian, *Level Set Methods: Evolving Interfaces in Geometry, Fluid Mechanics, Computer Vision and Materials Sciences*, 2nd ed. Cambridge University Press, University of California, Berkeley (1999).
16. D. Adalsteinsson and J. A. Sethian, "A fast level set method for propagating interfaces," *J. Comput. Phys.* **118**, 269–277 (1995).
17. J. A. Sethian, "A fast marching level set method for monotonically advancing fronts," *Proc. Natl. Acad. Sci. U.S.A.* **93**(4), 1591–1595 (1996).
18. R. Malladi and J. A. Sethian, "A real-time algorithm for medical shape recovery," in *Proceedings-of-the IEEE International Conference on Computer Vision (ICCV'98)*, pp. 304–310 (Jan. 1998).
19. N. Sochen, R. Kimmel, and R. Malladi, "A general framework for low level vision," *IEEE Trans. Image Process.* **7**(3), 310–318 (1998).
20. C. Ortiz de Solorzano, R. Malladi, S. A. Lelievre, and S. J. Lockett, "Segmentation of nuclei and cells using membrane-related protein markers," *J. Microsc.* **201**(3), 404 (2001).
21. W. Barrett and E. Mortensen, "Interactive live-wire boundary extraction," *Med. Image Anal* **1**(4), 331–341 (1997).

Automatic and Segmentation Based Registration of Serial Mammary Gland Sections

I. Arganda-Carreras[1,2], R. Fernandez-Gonzalez[2,3], C. Ortiz-de-Solorzano[2]
[1]Biocomputing Unit, National Centre of Biotechnology, Escuela Politecnica Superior, Universidad Autonoma de Madrid, Madrid, Spain
[2]Bioimaging Group, Life Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA
[3]UC Berkeley / UC San Francisco Joint Graduate Group in Bioengineering, Berkeley, CA, USA

Abstract—**We present two new methods for automatic registration of microscope images of consecutive tissue sections. Both methods are based on the images gradient correlation, but the first one makes use of the entire image information whereas the second one starts from the segmentation result. They represent two possibilities for the first step in the 3-D reconstruction of histological structures from serially sectioned tissue blocks. The aim lies in aligning the sections in order to place every relevant shape contained in each image in front of its corresponding shape in the following section. This is accomplished by finding the best rigid body transformation (translation and rotation) of the image being registered, by maximizing a matching function. To reduce computing time, we use a multiresolution pyramidal approach that seeks the best registration transformation in increasing resolution steps. In each step, a subsampled version of the images is used. The gradient of the images is computed using a Sobel operator. Then, the gradient image is binarized using an automatic threshold and the distance-transform of the binary image is computed. A proximity function is then calculated between the distance image of the image being registered and that of the reference image. The transformation providing a maximum of the proximity function is then used as the starting point of the following step. This is iterated until the error is below a minimum value.**

Keywords—**Automatic registration, Image processing, Biotechnology**

I. Introduction

A correct visualization of the morphology and functionality of the mammary gland, as similar as possible to the specimen living conditions, is basic in the study of normal mammary gland biology and its neoplastic variants (i.e. breast cancer). With this aim, we reconstruct mammary gland epithelial structures from fully sectioned paraffin tissue blocks, after it is stained with histological and protein or genetic markers. Our lab has developed a software microscopy system [1] that handles a microscope, scans sections of mammary gland tissues, and processes these images and reconstructs the morphology of the gland using the section information. As a starting point for the 3-D reconstruction of the histological structures, the system employs the section contours found by using our 2-D segmentation tool. However, as a first step towards an accurate 3-D reconstruction, a proper alignment of the images of sections is needed, in order to place every relevant 2-D structure in one section in front of its correspondent 2-D structure in the following section.

A perfect image alignment, also called registration, appears impossible due to human interaction in the section creation process (manual sectioning can cause non-linear distorting effects, such as deformations, folds, tears or cuts in the tissue), as well as to the natural differences between sections caused by the cut distance.

This first step of the registration process aims at achieving the best rigid body-transform for each pair of images for consecutive sections. This can be easily reached with a manual process, marking alternatively three points in every image and calculating the rigid body-transform (i.e. rotation plus translation) that minimizes the lineal quadratic error between every pair of points. However, manual registration of hundred of sections, as required when working with complete cases, could become very slow and tedious.

Although a perfect registration is not reachable, an accurate approach is usually enough, either for consolidating the tri-dimensional reconstruction, or for obtaining a base for following non-rigid registrations. Therefore, a rigid affine transformation will allow us to solve the problem of the registration. While this paper explains the solution for our mammary gland tissue sections alignment, it can be applied to many other different kind of images.

II. Methodology

Hematoxylin and Eosin were used to stain 5 micron serial sections from paraffin tissue blocks. The tissue was either normal mouse mammary gland or human tissue from biopsies diagnosed with ductal carcinoma in situ in the breast (DCIS). Every section image was taken in grayscale at low resolution (2.5X magnifications), and registered with the previous one in the case following the algorithm exposed in the next paragraphs.

As in all automatic registration methods, this system requires that the matching function addressing the algorithm measures the degree of similarity between the image to be aligned (or target image) with respect to the reference

This work has been supported by the US Department of Defense and California Breast Cancer Research Programs under grants DAMD17-00-1-306 and 8WB-0150 (respectively), by the LBNL Laboratory Directed Research and Development Program, under contract CSLD12 and by the National Institute of Health under grant 1R01HL67465-01. I. Arganda-Carreras is being supported by a predoctoral FPU-CAM fellowship since October 2003.

image, in order to achieve the best transformation. The criterion selected here is the *Hierarchical Chamfer Matching Algorithm* for registering [2].

In this method, a contour is projected in different positions on a distance image. The reference image is transformed in a distance image, meaning that the edges of the structures appearing in the image receive a value of white (the highest one) and the rest of the pixels present gradually smaller values depending on the distance to the edges (whiter as closer to the edges). Once the distance image is calculated, the system attempts to match it with the contour image, a binary image obtained by applying an automatic threshold function over the gradient of the image to be aligned. Before going on with the matching function, the contour image is transformed with three degrees of freedom: translations in the in x- and y- directions and rotation around the x-axis. Since the proximity function is calculated as the sum of the scalar product (pixel by pixel) of the contour, the distance images and the edges presenting white values, a perfect matching between images should then provide the biggest value as function result. Therefore, when this sum is maximized, the best transformation - and consequently alignment - is achieved.

Considering the huge size of the images being processed - usually around 6000x6000 pixels and 35 Megabytes each - the system will proceed over the highest resolution images. Therefore, the first step in the algorithm must be a subsampling of both target and reference images, which will produce an important reduction in the dimension of the images, and consequently in the computational time. The search for the local maxima starts in two reduced images generated from the original ones with as low a resolution as possible. The best match(ing) between the images applying all possible rotations and translations to the target will be calculated. Namely, the system uses all possible rotations from 0° to 350° in steps of 10° and all possible translations in different steps depending on the image size and guaranteeing at least 50% of image overlapping. This way this brute-force first level allows us to find an initial registration that can be used as a starting point for the following step, where the rank of translations and rotations will be reduced to the neighborhood of the previous result, and rotation and translation steps decreased in order to better refine the registration. So the problem is approximated step by step, and the result of every step is the starting point for the following one. That means that the algorithm presents a pyramidal organization, and uses different degrees of resolution and variation ranks in the rigid transformations at every pyramid level. Usually, the bottom of the pyramid is the less subsampled image (if the scale factor is 2, one pixel in level n corresponds to four pixels in level $n - 1$).

Since the system decreases gradually the subsampling factor in every level, lower levels involve more image information, providing more accurate results. A complete description of one level process is illustrated in Fig. 1.

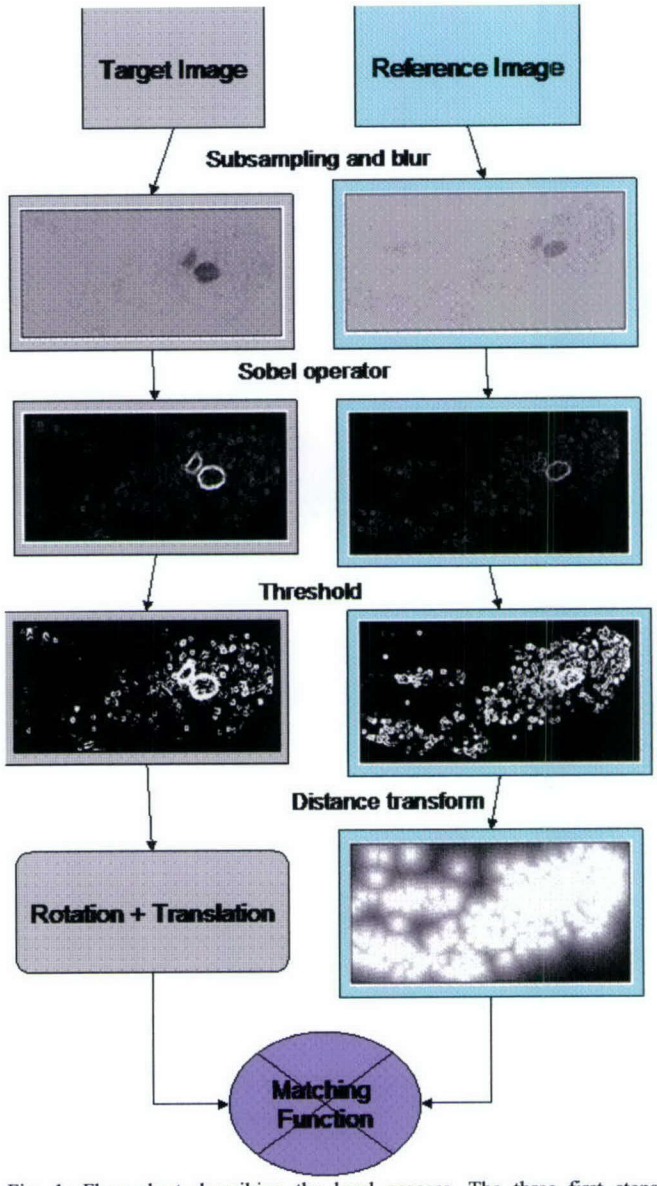


Fig. 1: Flow chart describing the level process. The three first steps (subsampling with blur, gradient and threshold) are equal for both images. At the fourth step the target image suffers the rigid body transformation and the reference is converted in a distance image. Same steps are applied in every level modifying only the subsampling parameters and the ranks in the affine transformation.

The second registration method proposed in this paper is a natural extension of the method already exposed. Taking advantage of the automatic segmentation tool developed in our lab [3], a new algorithm can be built starting from the segmentation results (Fig. 2). The segmentation produces in every image a list of contours, which can be reused in order to create a new binary image (black background and white contours). It seems obvious that this binary image could substitute the thresholded images appearing in the previous method, improving the final result, given the fact that the contours obtained in this segmentation process resemble

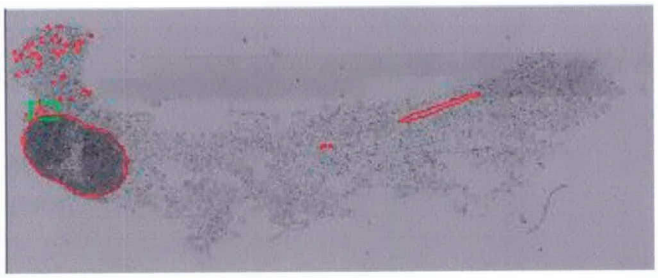


Fig. 2: Example of mice breast mammary gland section segmented automatically by our lab tool [3]. The contours in red represent the limits of relevant shapes, such as ducts, tumors or lymph nodes.

more accurately the section information needed by the expert.

Notice here that the frequent user of the segmentation tool is a mammary gland biologist, whose experience allows setting up the tool parameters in order to distinguish the noise from the relevant information, an experience that was not supplied to the previous automatic registration method (let us call it standard method). Another advantage in this second method (let us call it shape registration) is the reduction in the complexity of the image preprocess, i.e. only the two contour images need to be created (considering the subsampling factor of the correspondent pyramid level) and then the distance transform of the contour image corresponding to the reference image is calculated. Next, the normal rigid transformation and matching function will be applied through the different levels until getting to the optimum registration values (Fig. 3).

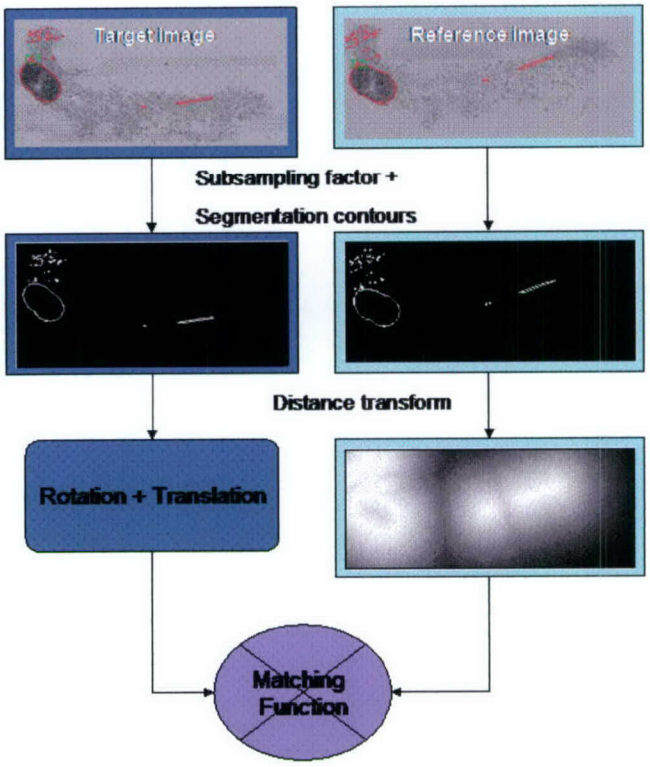


Fig. 3: flow chart describing one level of the registration process for the "shape registration" method. Extracting the segmentation information from the original images and rescaling the contours depending on the subsampling factor, the binary images are built. As before, the reference is transformed in a distance image and the target is rotated and translated before applying the matching function. For this example, same image as used in Fig. 2 was processed as reference image.

III. Results

Once the correct operation of the system was demonstrated experimentally for artificial cases, that is, cases with the same section rotated and translated, the method was applied to real mammary gland sections with fair results. One way of displaying the performance of the registration algorithm consists of creating a color image combining both target and reference images, representing the original reference image in red, and the target image, after being transformed according to the parameters obtained, in green. Overlapped structures should then appear in yellow, as it corresponds to the sum of green and red. Therefore, a perfectly aligned pair of images will present most of the areas in yellow and an incorrectly aligned pairs will show areas with intermediate colors (green towards red). Fig. 4 shows in this way the difference in accuracy between two consecutive system levels and its implicit error minimization.

Running the application on a PC Pentium IV (2.66 GHz, with 512 Megabytes of RAM memory) under Linux, the standard method of automatic registration on two typical sections (around 35 Megabytes each) is accomplished in 4 or 5 minutes using 3 different levels of resolution. Due to less process charge, the second method exposed or shape registration, is even faster than the previous one and only takes around 2 minutes for the same pair of sections.

Usually two pyramid levels achieve the optimum registration parameters. For more complex images, three levels are strongly recommended.

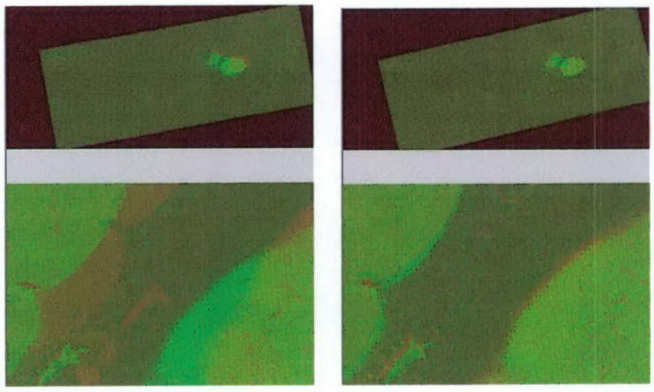


Fig. 4: Complete result and detailed zoom of the first and second level (respectively) in the standard automatic registration algorithm. Yellow areas mean structure overlapping and red and green areas represent incorrect alignment. These color pictures allow us to visualize the error reduction evolution.

IV. Discussion

As stated at the beginning of the paper, consecutive sections could present non-rigid deformations due to human interaction in the section elaboration process. Therefore, in these specific cases, an affine transformation could be insufficient for a perfect alignment and either local corrections in the rigid registration result or smooth filtering over the volume after the tri-dimensional reconstruction would then provide a more accurate approximation to the problem. Regardless, the rigid registration supplies a suitable result in most of our cases, and can be the first step towards a fully non-linear elastic registration.

The systematic search followed by our method, even if it is improved with a 50% of forced image overlapping and different rotations and translations steps depending on the image size and pyramid level, could be reviewed in order to find an optimization method that allows us to leave our current brute force method and decreases the computational time taken by the system.

V. Conclusion

We described a fully automatic algorithm for the registration of microscopy images of consecutive tissue sections, and a variant of the same method based on the previous segmentation of the images. The system is based on a multiresolution and pyramidal approach in order to calculate the optimum rigid transformation between Hematoxylin and Eosin (H&E) pairs of sections.

Acknowledgment

I. Arganda-Carreras thanks ImageJ open source project for provided code (http://rsb.info.nih.gov/ij/).

References

[1] Fernandez-Gonzalez R., Jones A., Garcia-Rodriguez E., Chen P.Y., Idica A. , Barcellos-Hoff M.H., Ortiz-de-Solorzano C. "A System for Combined Three-Dimensional Morphological and Molecular Analysis of Thick Tissue Specimens", *Microscope Research and Technique* 59(6): 522-530, 2002.

[2] R. Hult, *3-D reconstruction of insect ganglia*, Thesis work. Examensarbete Nr 22, Centre for Image Analysis, Uppsala, 1995.

[3] Fernandez-Gonzalez R., T. Deschamps, Idica A., Malladi R., Ortiz de Solorzano C. *Automatic segmentation of histological structures in mammary gland tissue sections*. Journal of Biomedical Optics (accepted, scheduled for publication in May 2004).

Appendix 4

A Tool for the Quantitative Spatial Analysis of Mammary Gland Epithelium

Rodrigo Fernandez-Gonzalez[1, 2], Carlos Ortiz de Solorzano[1]
[1]Life Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA
[2]Joint Graduate Group in Bioengineering, UC Berkeley/UC San Francisco, CA, USA

Abstract—**There are many problems in mammary gland biology where the spatial distribution of the cells may be playing a fundamental role. Thus, a tool that can quantitatively measure that distribution would be extremely helpful in order to solve some of the previously mentioned problems. Here we present a method for the spatial analysis of mammary epithelium based on a multiscale study of neighborhood relationships. A function to measure those relationships, *M*, is introduced. The refined Relative Neighborhood Graph is then presented as a method to establish vicinity relationships between epithelial cell nuclei in the mammary gland. Finally, the method is illustrated with two examples that show interactions within one population of epithelial cells and between two different populations.**

Keywords—**Mammary gland, quantitative, spatial distribution.**

I. Introduction

The spatial distribution of cells within epithelial tissues plays a fundamental role in development, function and regeneration of these tissues [1, 2]. For example, in mammary gland development, cap cells with invasive properties line the surface of the growing ducts, which extend through the fat pad that embeds the gland. In the alveolar units that cap the ducts of a mature gland, secretory epithelial cells line the lumen of the ducts. After pregnancy, this mature luminal epithelium secretes milk into the ducts. Myoepithelial cells are arranged around the ductal tree, as well as around the terminal alveolar units. Upon, the appropriate stimulus they contract, thus forcing the milk through the ducts towards the nipple [3].

Many of these spatial phenomena have been previously described in a qualitative way but, due to the lack of appropriate tools, most of them have not been quantitatively studied. For example, the colocation in the patterns of expression of the estrogen (ER) and the progesterone (PR) receptors in luminal epithelial cells [4]; the fact that proliferation markers are not expressed by ER^+ cells [5]; or the possible presence of a niche -a highly organized pattern of different cell types- around mammary stem cells [6] have never been assessed from a quantitative, spatial point of view.

In order to address these questions, we have developed a quantitative spatial analysis tool that we have integrated into our 3D microscopy system [7]. In this paper we introduce that tool and show two examples obtained on real data, thus demonstrating how we will use it to address some of the questions mentioned above.

R Fernandez-Gonzalez was supported by a predoctoral fellowship from the Department of Defense Breast Cancer Research Program (BC020294).

II. Methodology

A. Tissue processing

Mammary gland tissue blocks are sectioned at 5 μm, and the sections are immunostained for the appropriate antigens. A counterstain is used that allows us to study the morphology of the tissue (e.g.: DAPI). Low magnification (2.5X) images of the counterstaining of all the sections are then automatically acquired using a motorized Zeiss Axioplan I microscope coupled with a monochrome XilliX Microimager CCD camera. This is done automatically by scanning the area of the slide occupied by the tissue and tiling together all the individual snapshots into a single image of the entire section.

The next step consists of automatically annotating the structures of interest (ducts, tumors, ...) in these low magnification images [8]. These annotated structures are then used to reconstruct a three-dimensional model of the sample. With this model we can track the morphology of the tissue to determine which are the areas where a spatial analysis might be more interesting. After selecting these areas, the system asks the user to place the right fluorescent section(-s) on the stage, and high magnification (40X) images of the immunostaining of the chosen areas are acquired. In these images nuclei are manually annotated with dots and visually classified, forming a point pattern; they can also be automatically segmented and quantified [9]. In the later case, the center of mass of each nucleus is computed to obtain a point pattern of nuclei markings.

B. M function analysis

B.1. Definitions

Given a set of points $\{n_1, ..., n_{Nc}\}$ representing the nuclei belonging to population C in the study area, we define n_{iCr}, the number of neighbors of nucleus n_i belonging to population C within distance r; n_{ir}, the total number of neighbors of n_i (belonging to any population) within distance r; N_C, the total number of nuclei belonging to population C within the area under study; and N, the total number of nuclei in that same area (Fig. 1).

B.2. Single-variable analysis

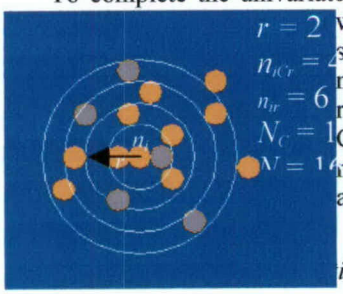
The distribution of epithelial cells across mammary tissue is not homogeneous: they can only be located at ducts, end buds and alveolar structures, but never in the fat pad surrounding the gland, which also shows up in the images. For that reason, we cannot do our spatial analysis

with Ripley's K function [10], which has traditionally been used in other fields for this task. We now assume a space with heterogeneous nuclei density. In this space, the total number of nuclei on an area measures the size of the set of possible locations for an epithelial nucleus in that area: any of those points could be occupied by a nucleus. Thus, we can measure the density of nuclei belonging to population C as a ratio of nuclei numbers, and so, we define [11]:

$$M(r, C) = \frac{\sum_{i=1}^{N_c} \frac{n_{iCr}}{n_{ir}}}{N_C} / \frac{N_C}{N} \quad (1)$$

The numerator of (1) computes the average density of neighbors belonging to population C within distance r, and then compares that value to a benchmark: the density of nuclei belonging to population C in the entire study area. Therefore, clustered patterns of nuclei will have $M(r, C) > 1$, with a peak at the cluster size. On the other hand, regular patterns will have $M(r, C) < 1$. Finally, random distributions will have $M(r, C) = 1$. In general, we can say that $M(r, C) = k$ implies that the density of nuclei belonging to population C within distance r is k times that of the entire area under study.

To complete the univariate analysis we need to have a way to establish the significance of our measurements. For that reason, we run m Monte Carlo simulations of the nuclei distribution within the area of interest. The



ion examples.

simulations are set up by preserving the nuclei locations and randomly assigning the population where each nucleus belongs. We compute the M function for each one of these simulations ($M_i(r, C)$, i = 1, ..., m) and calculate $U(r, C)$ and $L(r, C)$:

$$U(r, C) = \max_{i=1,\ldots,m} (M_i(r, C)) \quad (2)$$

$$L(r, C) = \min_{i=1,\ldots,m} (M_i(r, C)) \quad (3)$$

Now we can plot $M(r, C)$, $U(r, C)$ and $L(r, C)$ in the same graph. Peaks of $M(r, C)$ above $U(r, C)$ are evidence of significant clustering (with confidence level $\alpha = 1/(m + 1)$). Similarly, troughs below $L(r, C)$ represent significant regularity or dispersion. Any nuclei distribution with no significant peaks or troughs can be considered to be random.

B.3. Multiple-variable analysis

The M function analysis described in the previous section can be used to study the distribution of cells within a single population. However, most of the problems introduced in section I involve two or more cell populations. In order to study this type of problems, we can modify (1) to get:

$$M(r, C_1, C_2) = \frac{\sum_{i=1}^{N_{C_1}} \frac{n_{iC_2r}}{n_{ir}}}{N_{C_1}} / \frac{N_{C_2}}{N} \quad (4)$$

where C_1 and C_2 are the two populations under study, N_{C1} and N_{C2} are the number of nuclei in each one of those populations and n_{iC2r} is the number of nuclei belonging to population C_2 within distance r of nucleus n_i (with n_i belonging to C_1). It is easy to see how these equation could be extended to three or more populations.

Now M values larger than 1 are indicative of attraction between populations C_1 and C_2 (a special case of this is *colocation*, i.e., attraction at distance $r = 0$); values smaller than 1 indicate repulsion; and $M(r, C_1, C_2) = 1$ shows independence of the spatial distributions of both populations at distance r. However, significant values of $M(r, C_1, C_2)$ maybe due either to actual interactions between both populations or to the patterns of each one of them. For this reason, we set up our Monte Carlo simulations preserving the locations of the nuclei belonging to population C_1 and redistributing the location of population C_2. Thus, we control for the C_1 pattern. The same process is applied to $M(r, C_2, C_1)$. Finally, significant interaction at distance r is only accepted if both $M(r, C_1, C_2)$ and $M(r, C_2, C_1)$ are significantly different from randomness.

C. Refined RNG

In the previous section we have used the shortest Euclidean distance to measure how far apart two nuclei are. However, this is not the best way to represent vicinity. In fact, the shortest Euclidean distance is often obtained through luminal areas where nuclei cannot be located. This is in contrast with cell-to-cell signaling in the epithelium, which normally occurs through intermediate cells [12]. Therefore, we decided to model our tissue using a graph where the nodes are the different nuclei, edges represent neighborhood relationships and distances can be measured as the number of edges between two nuclei.

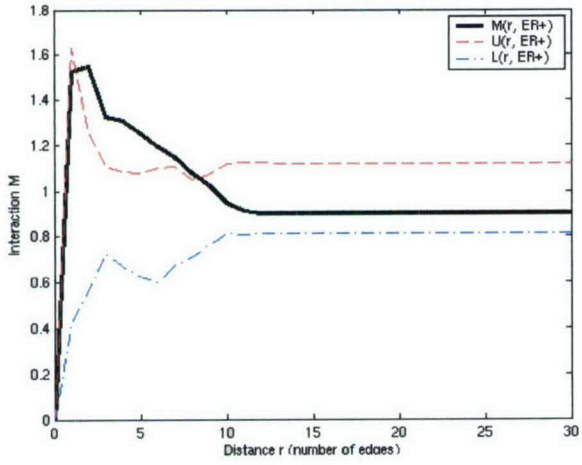
We start out by building a Delaunay triangulation using the nuclei markings as nodes. This provides a preliminary tessellation where we can already measure distances as number of edges. On top of this triangulation we can now build a Relative Neighborhood Graph (RNG). Here, we preserve an edge if an only if the two nuclei on its sides (n_i and n_j) are relatively close [13], that is:

$$d(n_i, n_j) \leq \max(d(n_i, n_k), d(n_j, n_k)) \quad (5)$$
$$\forall\ k = 1, \ldots, N, k \neq i, j$$

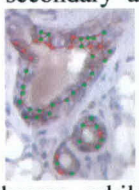
where $d(n_i, n_j)$ is the length of the edge between n_i and n_j. In other words, what this definition states is that an edge in the Delaunay triangulation is preserved if the nuclei on its sides are at least as close to each other as they are to any other nucleus in the graph. With this we obtain the RNG. Finally, we do a refinement step where we get rid of all the edges which are too large and we force connections between nuclei which are too close (using the shortest Euclidean distance) to not to be neighbors. Now, using Floyd's or Dijkstra's algorithms [14], we can easily build a table with the shortest distance (measured as the number of edges) between each pair of nuclei in the graph.

III. Results

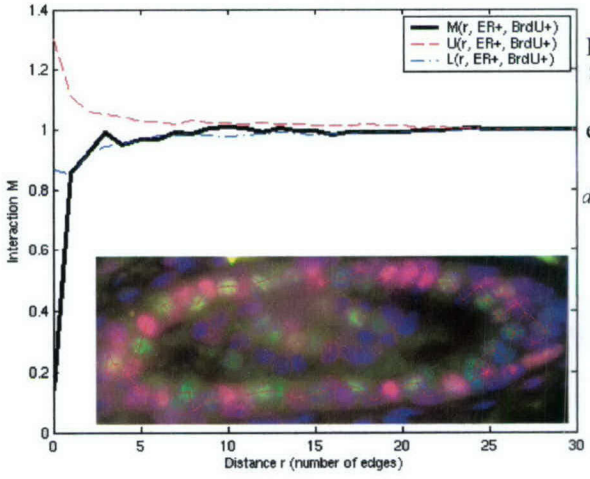
We run our spatial analysis tool on a set of synthetic images to test for its accuracy at detecting interactions both within and between populations. Then we went on to the analysis of real tissue samples. In this section we describe two different examples. For the first one the tissue was obtained from a transgenic mouse overexpressing the HER2 gene, a growth factor receptor whose human counterpart is overexpressed in about 30% of breast cancers. Sections were taken from this sample and immunostained for HER2 using diaminobenzidine (brown precipitate). The nuclei were counterstained with hematoxylin (blue). High magnification images of certain areas in these sections were acquired, and the nuclei in those areas were manually annotated. Fig. 2 shows the spatial analysis of the $HER2^+$ population in one of those areas (inset). Positive nuclei are marked with red dots, negative ones are green. The analysis indicates the presence of clustering at small distances around the nuclei (at 2 to 8 nuclei of distance as measured by edges in the refined RNG), with a peak at distance $r = 2$. This peak reveals the presence of clusters of $HER2^+$ cells with a radius of 2 nuclei and a density of $M = 1.55$.



For the second example we obtained tissue from a wild type mouse which had been given a constant dose of BromodeoxyUridine (BrdU) for two weeks. BrdU is a thymidine analog that gets incorporated into the DNA of the cells that undergo mitosis. The tissue was sectioned, and double immunofluorescence staining was carried out on the sections. BrdU was detected using a secondary antibody



labeled with Alexa 568, a red fluorochrome, while Alexa 488 (green) was used to detect ER^+ cells. Nuclei were counterstained with DAPI (blue). Once again, high magnification images of some areas were taken, and nuclei were manually annotated. Then, multiple-variable analysis of the interaction between ER^+ and $BrdU^+$ cells was carried out. Fig. 3 shows an example of this type of analysis ($M(r, ER^+, BrdU^+)$). Nuclei have been removed from the inset image for clarity. The graph for $M(r, BrdU^+, ER^+)$ is very similar to the one shown here. Thus, there seems to be repulsion between both populations at very small scales. Actually, the peak of this repulsive interaction is at distance $r = 0$ ($M = 0.15$), i.e., ER^+ and $BrdU^+$ cells do not colocalize most of the times, but do colocalize occasionally. This result, whose intensity and extent we can now quantify using the M values, has previously been described in the literature.



IV. Discussion

Describing spatial phenomena is of great importance in biology in general, and particularly in mammary gland studies. In order to do this in a way that provides consistency and high

throughput, quantitative tools for the spatial analysis of samples are required. In this paper we have presented a method that automatically provides a measurement of the way cells interact within one population, as well as of the different types of interaction that might occur between the different cell populations present in a tissue.

Our approach is based on a multiscale analysis of the number of neighbors belonging to the population under study, followed by comparison to a benchmark, the total density of nuclei within that population in the entire study area. Thus, we define the *M* function, which takes into account the heterogeneous distribution of the epithelium within mammary tissue. This function, -together with the analysis scheme where it is embedded-, allows for unsupervised analysis of large data sets in a reasonable time, since it does not include any complex calculation. The method provides comparability of concentration measurements across populations; remains unbiased concerning different scales; and can be modified depending on the desired significance level.

In order to define neighborhood in a way that takes into account the histology of the tissue as well as differences in cell size/image magnification, we create a refined RNG that has the nuclei markings as its nodes. The connections in this graph represent vicinity in a way that faithfully depicts what nuclei might be directly interacting with each other in the tissue.

In the near future we are planning on using this tool to address several problems, including colocation/interaction studies of ER^+, PR^+ and $HER2^+$ populations in both wild type and transgenic mice, or characterization of the distribution of label-retaining cells (a population of cells likely to be enriched for mammary stem cells) with respect to other populations present in the mammary epithelium, thus trying to unveil the presence of a niche around stem cells similar to the ones observed in other organs. Since both of these problems are inherently three-dimensional, we are currently working on adding one more dimension to our analysis scheme. This extended functionality, together with the implementation of methods taken from the fields of pattern recognition and mathematical morphology on graphs, should help us explore in further detail the possible determinants of interaction both within and between cell populations.

V. Conclusion

In this paper we have presented a method for the spatial analysis of mammary epithelium. Thus, we have created a tool to quantitatively measure what previously could only be qualitatively described. Our multiscale method is consistent, comparable across populations and allows for automatic, high-throughput analysis of large data sets. The use of this approach to study problems where interactions between cells are expected will greatly contribute to the detailed description of these phenomena.

References

[1] E. Fuchs, "Beauty is skin deep: the fascinating biology of the epidermis and its appendages", *Harvey Lect.* 94, pp. 47-77, 1998.

[2] W. Imagawa, J. Yang, R. Guzman, S. Nandi, "Control of mammary gland development" in *The Physiology of Reproduction 2nd Ed.*, E. Knobil, J.D. Neill Eds. New York Raven Press., 1994, ch. 3, pp. 1033–1063.

[3] M. Smalley and A. Ashworth, "Stem cells and breast cancer: a field in transit", *Nat Rev. Cancer*, vol 3, no.11, pp. 832-844. Nov. 2003.

[4] G. Shyamla, Y.C. Chou, S.G. Louie, R.C. Guzman, G.H. Smith, S. Nandi. " Cellular expression of estrogen and progesterone receptors in mammary glands: regulation by hormones, development and aging", *J. Steroid Biochem. Mol. Biol.*, vol. 80, no. 2, pp. 137-148, Feb. 2002.

[5] R.B. Clarke, A. Howell, C.S. Potten, E. Anderson, " P27(KIP1) expression indicates that steroid receptor-positive cells are a non-proliferating, differentiated subpopulation of the normal human breast epithelium", *Eur. J. Cancer*, vol. 36, suppl. 4, pp. 28-29, Sep. 2000.

[6] N.J. Kenney, G.H. Smith, E. Lawrence, J.C. Barrett, D.S. Salomon, "Identification of Stem Cell Units in the Terminal End Bud and Duct of the Mouse Mammary Gland", *J. Biomed. Biotechnol.*, vol. 1, no. 3, pp. 133-143, 2001.

[7] R. Fernandez-Gonzalez, A. Jones, E. Garcia-Rodriguez, P.Y. Chen, A. Idica, S.J. Lockett, M.H. Barcellos-Hoff, C. Ortiz-De-Solorzano, "A system for combined three-dimensional morphological and molecular analysis of thick tissue specimens", *Microsc. Res. Tech.*, vol. 59, no. 6, pp. 522-530, Dec. 2002.

[8] R. Fernandez-Gonzalez, T. Deschamps, A. Idica, R. Malladi, C. Ortiz-De-Solorzano, "Automatic segmentation of histological structures in mammary gland tissue sections", *J.Biomed. Optics*, in press.

[9] N. Malpica, C. Ortiz-De-Solorzano, J.J. Vaquero, A. Santos, I. Vallcorba, J.M. Garcia-Sagredo, F. del Pozo, "Applying watershed algorithms to the segmentation of clustered nuclei : defining strategies for nuclei and background marking", *Cytometry*, no. 28, pp. 289-297, 1997.

[10] T.C. Bailey and A.C. Gatrell, *Interactive Spatial Data Analysis*. Essex, England: Prentice Hall, 1995, pp. 90-95, 103-105, 117-131.

[11] E. Marcon and F. Puech, "Measures of the geographic concentration of industries: improving distance-based methods", *Cahiers de la MSE*, 18, p. 22, 2003.

[12] B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts and P. Walters, *Molecular Biology of the Cell, 4th ed.*. New York, NY: Garland Science, 2002. Ch. 15.

[13] G.T. Toussaint, "The relative neighborhood graph of a finite planar set", *Pattern Recognition*, vol. 12, pp. 261-268, 1980.

[14] B.R. Preiss, *Data structures and algorithms with object oriented design patterns in Java*, John Wiley & Sons, 1999. Available: http://www.brpreiss.com/books/opus5/

Appendix 5

Three-dimensional Histology of the Mammary Gland: An application to the study of hormone receptor expression during normal mammary gland development

Ouahiba Laribi, Abbey Hartland, Rodrigo Fernández-González, Ignacio Arganda-Carreras, Adam Idica, Mary Helen Barcellos-Hoff, Carlos Ortiz-de-Solórzano

Life Sciences Division, Lawrence Berkeley National Laboratory
University of California
USA

Corresponding author:
Carlos Ortiz de Solórzano, Ph.D.
Lawrence Berkeley National Laboratory
1 Cyclotron Road
Berkeley, CA 94720
USA
Email: CODeSolorzano@lbl.gov
Phone: (510) 486 4923

Keywords: mammary gland development, hormone receptors, estrogen receptor α, progesterone receptor, three dimensional microscopy, image analysis, scientific visualization.

ABSTRACT

Reproductive hormones, acting through their cognate cellular receptors are known to play an important role in normal mammary gland development and neoplasia. Although a great deal of information exists about hormone levels during the different developmental and menstrual stages, less information exists about the expression of cellular hormone receptors.

In this paper we have quantified the expression of estrogen receptor α (ERα) and progesterone receptor (PR) using a 3D computer-assisted microscopy system that can reconstruct tissue blocks from multiple 2D tissue sections and integrate morphological and molecular information from sequential histological and immunostained sections.

Using this system we have reconstructed the mammary ductal tree of female mouse at different development stages (puberty, adulthood and aging) and quantified the expression of the receptors in the entire mammary gland, to show the differences in number of receptor expressing cells between different morphological structures and/or developmental stages.

INTRODUCTION

Histology –the microscopic study of tissue structure- is the primary source of information about complex biological systems. The histo-pathologist determines the status –developmental, pathological- of an organ, by taking a combined look at the morphology –both at the tissue and cellular level- and the molecular make-up of *ex vivo* tissue samples. The tools available to the histo-pathologist are constantly improved by the continuous influx of novel molecular markers that, coupled to a precise morphological assessment, help determining what is considered normal or abnormal (pathological) tissue behavior.

Very much like it was practiced centuries ago, histo-pathology involves visual inspection of two dimensional –flat- stained tissue sections. In one's daily routine, the histo-pathologist observes particular morphological and molecular events within the cells or their surroundings, in the local 'geography' of the section. In cancer, for instance, the number and/or distribution of morphologically abnormal cells, combined with the presence of markers for carcinogenic transformation (e.g. overexpression of oncogenes, genetic aberrations) and with the overall distortion of tissue architecture is used to diagnose the disease and predict its progression.

While traditional histo-pathology remains the gold standard for the analysis of two-dimensional tissue sections, it largely ignores the three-dimensional information present in tissues except for sketchy 3D impressions mentally integrated by the histo-pathologist upon looking at a few consecutive sections, or by looking at very low-resolution tissue whole mounts. Thus it is reasonable to postulate that a full three-dimensional approach, that provide the histo-pathologist with an accurate 3D virtual rendition of tissue structures, would dramatically improve his or her understanding of the tissue, and thus greatly facilitate its pathological assessment. Furthermore, correlating the 3D tissue reconstruction with quantification of molecular markers at cellular level in regions of the tissue, selected precisely by their morphology, can produce vital information for uncovering highly heterogeneous tissue processes, both normal and pathological.

Understanding the complexity of the architecture of an organ and its heterogeneity requires a three-dimensional, multi modal approach that would have been unthinkable a decade ago. However, the steady increase in computational power and storage capacity of computers, coupled to the development of highly sensitive, inexpensive imaging devices, motorized microscopes and new and improved image processing techniques have placed this goal within our reach. However, as far as we know, there are very few studies to this date that make use of computers to address this important issue (Ohuchi et al 1994, Ohtake et al 1995, Ohtake et al 2001).

Our laboratory has developed R3D2, a computer assisted microscopy system that consists of an integrated environment for the semiautomatic acquisition, annotation and reconstruction of tissue structures from fully sectioned thick tissue specimens (Fernández-González et al 2002). R3D2 can be used to reconstruct epithelial structures, as defined by standard histological staining (H&E) of thin tissue sections, and then map immunostained proteins and genes at cellular level in the three-dimensional reconstruction of the tissue. We refer to (Fernández-González et al 2002) for a full technical description of R3D2.

In this paper, we have used R3D2 to study the three-dimensional distribution of hormone receptors in the mouse mammary gland, during the principal phases of its development.

The study of the expression of hormone receptors has special interest due to their relevance in the development of the gland and their known role in cancer. The action of estrogen is primarily regulated through its cognate receptor protein, estrogen receptor (ER). ER is heterogeneously expressed throughout the epithelium and stroma, and exists in two isoforms: ER-α and ER-β. While ER-α is amplified in over half of all breast cancers and perhaps contributes to aberrant growth, ER-β has been shown to have an antiproliferative effect on cancer cells (Lazannec G. et al. 2001). Though these receptors often coexist within an epithelial cell nucleus, each is responsible for a distinct physiological activity in response to estrogen binding (Levin E.R. 2001, Cheng G. et al. 2004). Progesterone is regulated through a similar nuclear receptor protein,

progesterone receptor (PR), which also exists in two forms: PR-A and PR-B. Unlike ER, PR expression is restricted to epithelial nuclei and is undetectable in surrounding stroma (Shyamala G. et al. 1997). Recent studies using mice knockout models have shown that estrogen signaling through ER-α is responsible for ductal morphogenesis (Boccinfuso W.P. et al. 1997), while progesterone signaling, though either isoform of PR, is responsible for lateral side branching and the maturation of the lobuloalveolar structures at the terminus of each duct (Shyamala G. 1999, Lydon J.P. 1995). Because ER-α and PR play such important roles in the normal development of mammary epithelium and in its neoplastic variants, we have chosen to study their heterogeneous patterns of distribution within the mammary gland at different stages of the life spam of our animal model, FVB mice.

METHODS

Animals and tissue collection

We used 10 wild-type FVB nulliparous female mice. Animal breeding, care and euthanasia were done following a protocol approved by the local animal welfare and research (AWRC) committee. Namely, two animals were sacrificed by CO2 inhalation followed by cervical dislocation at 6, 12, 18, 32 and 48 weeks of age. . Upon its death, the left inguinal (4L) gland was surgically extracted. The gland was fixed in 10% formaldehyde overnight and stained for whole mount.

Histology

The whole mounts were macroscopically observed and imaged with a digital camera. Both the entire fat pad and the area of the fat pad occupied by epithelium were manually delineated in the images and the percent of fat pad filled calculated as the ratio between the areas –number of pixels- inside both curves. After gross observation, the glands were embedded in paraffin and cut into 5μm-thick sections using a microtome. Each odd-numbered section was stained with hematoxylin and eosin (H&E) and mounted on a silane-coated glass slide.

Antibodies

The antibodies used were: Anti-ER-α, mouse monoclonal antibody 6F11 (Novocastra Laboratories), Anti-PR, rabbit monoclonal antibody (gift of Dr. Shyamala, Lawrence Berkeley National Laboratory), biotinylated anti-rabbit IgG (1:250, Vector Laboratories), and biotinylated anti-mouse IgG (1:250, Vector), Alexa 488 anti-mouse IgG (1:100, Molecular Probes) and Alexa 568 anti-rabbit IgG (1:100, Molecular Probes).

Immunohistochemistry

Each even-numbered slide was antibody-stained for the presence of ER-α or PR. Slides were warmed at 55°C for 30 minutes, washed in three baths of Histo-Clear (National Diagnostics) and rehydrated in a series of graded alcohol concentrations. Antigen retrieval was performed by microwaving slides in antigen unmasking solution (Vector Laboratories) for 3X 7 minutes.

Single Immunostaining

Peroxidase suppressor (Pierce) was used for 1 hour to quench endogenous peroxidase. All non-specific binding sites within the tissue were blocked with one hour of Super Block Blocking Buffer (Pierce) at room temperature (RT). Slides were incubated overnight at 4°C with the primary antibody (Anti-ER-α, mouse monoclonal antibody 6F11 (1:100) Anti-PR, rabbit monoclonal antibody (1:1800) and one hour at room temperature with the secondary antibody (biotinylated anti-rabbit IgG (1:250), and biotinylated anti-mouse IgG (1:250). Vectastain Elite ABC Kit (Vector) was applied for 35 minutes and revelation was done using DAB Substrate Kit for 10 minutes (Vector Laboratories). Staining was counterstained with Gill's formula hematoxylin and tissue was prepared for storage in Permount (Fisher Scientific).

Double Immunostaining

Tissue sections were incubated for one hour at RT with Super Block Blocking Buffer (Pierce) and overnight at 4°C with the first primary antibody (Anti-ER-α, mouse monoclonal antibody, 1:40) followed by one hour at room temperature with the first secondary antibody (Alexa 488 anti mouse, 1:100). Blocking Buffer was applied a second time for one hour and the slides were incubated with the second primary antibody (Anti-PR, rabbit monoclonal antibody, 1:900) overnight at 4°C and one hour with the second secondary antibody (Alexa 568 anti rabbit, 1:100). The slides were counterstained with 6-diamidino-2-phenylindole dihydrochloride (DAPI 1:1000, Sigma Aldrich) for 3 minutes and mounted with Vectashield (Vector Laboratories).

Tissue imaging

Accurate rendition of the architecture of the mammary gland requires the identification and tracking of all epithelial structures in serial sections followed by computer-based rendering. Using R3D2, our computer assisted microscopy system (Fernández-González et al 2002), we acquired low resolution image scans of all the H&E and immunostained sections. R3D2 controls a fully automated Zeiss Axioplan microscope connected to a CCD high-resolution black and white camera. To create the images, R3D2 automatically scanned the sections, moving the stage and tiling together single snapshots of the camera to create a mosaic-like wide view of the entire section. The Z (focusing) position of the microscope was automatically corrected when needed to prevent defocusing caused to inclination of the microscope stage, imperfect flatness of the slide, tissue folding, etc. Due to the non-uniform illumination of each field-of-view, the images showed a periodic pattern of background intensity that we corrected before the analysis (figure 1). For the low magnification scans we used a 10X 0.5 NA Zeiss Fluor objective, which provided acceptably low spherical aberration of the field of views and enough optical resolution to resolve all relevant epithelial structures in the images. However, due to the size of the sections at full resolution, the images were subsampled by a factor of 4 in both X and Y directions, for an effective 2.5X (25 times when considering the magnification of

the microscope tube) magnification. All the images of the H&E and immunostained sections were saved in ICS format (Dean P. et al. 1990) and stored in what we call "*cases*". R3D2 allows the user to browse the sections of the case, and access the registration, annotation, analysis and revisiting tools described below.

The number of sections per case ranged from 50 to 200 sections, each section using between 20 and 100MB of memory. Due to the large image size, which prevents from efficiently loading them into the memory of the computer, the image files were mapped into the memory, thus actually loading in the memory the parts of the images being displayed at full resolution.

Tissue reconstruction

A topologically accurate rendition of the epithelial structures (e.g. ducts, lobules, end buds) of a case requires that all the images of the sections be correctly registered (i.e. aligned), to ensure the continuity of structures crossing several sections. R3D2 provides an automatic registration tool that calculates the best matching between each pair of sections after rigid-body transforming one of them with respect to the other. This linear rigid-body transformation (i.e. rotation + translation) is accurate enough for a correct global alignment of the sections and therefore to create a topologically correct rendition of the structures. However, the manual sectioning process introduces non-linear effects (tissue tearing, stretching, folding, local artifacts, etc), which are not addressed in this study, where we are not as interested in the smoothness of the final reconstruction as we are in its topological accuracy. Thus we used R3D2's linear rigid body transformation tool to automatically align all the sections of the case. The algorithm works in batch mode for the entire case, and does not require any human interaction.

After registration, selected epithelial areas (ducts, TEBs, alveoli, lymph nodes) were manually delineated on the images of the H&E stained sections (figure 2) and parts belonging to the same structures that lie in different sections were manually joined to set the continuity of the structures. This process of grouping

identifies each object as a 3D volume, which can then be rendered in 3D using R3D2's triangulation algorithm.

Tissue revisiting

The work of the pathologist requires revisiting areas of the tissue at different magnification and relating areas from H&E stained and immunostained sections. When more than one slide is involved, this requires reloading the corresponding slide(s) and looking for the areas that want to be revisited. Given the complexity of the tissue and the difference in magnification, which often requires switching back and forth between a dry and an oil-immersion objective lens, this can be rather cumbersome and time consuming.

Our system preserves the spatial correspondence between the pixels of the low magnification images and the actual x/y coordinates of slides in the microscope stage. This greatly simplifies the process of revisiting the sections, which only requires clicking on the areas of the low magnification images that need to be revisited. R3D2 then calculates the appropriate microscope movement that places it in the desired area(s). This can be used to visually revisit the tissue or to acquire new images at higher/magnification (see below) or using a different color filter for multi-color image acquisition. This way, since the areas are identified at the computer on the acquired images and not on the actual slides, no switching between lenses is required, very much streamlining the revisiting process.

Using this feature, we revisited at high magnification (40X, Plan Neo, 1.39 NA) all the areas selected for rendering on the H&E sections. However, instead of visiting the H&E sections, we revisit the contiguous –properly registered- immunostained sections, to image the expression of the hormone receptors (ER and PR) in the epithelium of the selected areas.

All areas were imaged in color by consecutively imaging three black and white images after inserting the appropriate color filter in the light path of the microscope. To avoid manual interaction we used a crystal tunable RGB filter controlled by R3D2. When the areas were larger than a field of view of the

microscope at 40X, image scans were taken, the way it was done for the low magnification images, although in color.

All the images of the areas (in the order of 100, depending on the number of sections of the case) were stored in the *case*, related to their corresponding sections and selected structures. The images can be easily displayed by clicking in the corresponding areas of the low-resolution images. This action opens a new window that displays the high-resolution image and provides access to the analysis tools described below.

Statistical Analysis

R3D2's image analysis tools were used to quantify the presence and distribution of molecular markers in all the high magnification images of the areas of the case. Based on DAB staining, the luminal cells stained for a nuclear receptor (ER-α or PR) were interactively classified as negative or positive (Figure 3). If the areas contained more than one type of structure, masks were used to classify them separately. This way, all the epithelial fragments of the areas were classified under one of the following categories: large (collecting) duct, small duct (including alveolar areas and terminal ducts) or end bud. The classification as large or small duct was done considering the width of the duct, the amount of surrounding fibrous stroma and the location within the fat pad. After masking, the numbers of positive cells for each receptor was calculated for each type of structure separately.

RESULTS

Macroscopic Observation

Initial observation of the whole-mount glands revealed the fat pad to be 55-60% filled at 6 weeks. Several big terminal end buds (TEBs) (9 to 12) were detectable in the whole mount macroscopic images indicating intense ductal growth (figure 4). By 12 weeks the fat pad was completely filled and no terminal end buds were

detectable. At this stage, the epithelial tree was very dense with thick ducts and extensive lateral branching. The morphology remained the same at 18, 24 and 30 weeks. By 36 weeks of age, the gland started to regress, i.e., as shown by dramatically decreased lateral branching and alveoli structures (figure 5).

Microscopic Observation

Microscopic observation of the H&E tissue allowed us to see more detailed morphological changes. At high resolution, terminal end buds could be seen at 12 weeks, however, compared with those of the 6 weeks, their size was greatly reduced (figure 6). In addition, alveoli remained undeveloped until 18 weeks.

Microscopic observation of the immunostained tissue (figure 3) showed that the distribution of the receptors ER-α and PR agreed with previous studies (Shyamala G. et al. 20002). Both ER-α and PR were expressed in the luminal epithelial cells of the glands. Although ER-α was mainly expressed in the epithelium, it was also found in the fatty stroma and the fibrous stroma surrounding the epithelium of primary ducts. Neither ER-α nor PR was found in the myoepithelial cells that lay within the basement membrane or in the cap cells present on the tips of the TEBs.

3D Reconstruction

R3D2 allowed us to reconstruct the entire ductal tree of a mammary gland of a 6 weeks old animal (figure 7) and all the areas selected in all the other animals. In figure 7, all volumes could be selected individually (figure 8) and analyzed separately giving us information about the complex heterogeneity of the gland.

Molecular Distribution of the Hormone Receptors ER α and PR

The overall numerical analysis (figure 9) showed that both receptors were highly expressed in the 6-week animals, with 39% of cells expressing ER and 30% expressing PR. ER expression dropped to 28% in the 12-week animals, while PR remained at peripubertal levels. As the gland reached morphological maturity at

18 weeks, receptor expression dropped to 16% for both proteins, and remained low throughout the remainder of the 48 weeks.

Using the 3D reconstruction as described earlier, several structures of interest were tracked and classified as small ducts, large ducts or terminal end buds. We analyzed the cell-by-cell receptor status of each type of structure separately. Both receptors were consistently lower in large ducts than in small ducts, dropping to as low as 13% (figure 10). ER was highly expressed in 6-week small ducts (43%) and TEBs (42%), where most ductal elongation occurs. PR was also highly expressed in 6-week small ducts (38%), areas with copious lateral branching. No changes in the molecular distribution were observed after 18 weeks, although if the intensity of staining is an index of the levels of receptor expression, ER-α expression decreases gradually between 18 and 48 weeks. No changes were observed for PR expression levels (figure 5).

Co-localization of ER-α and PR

We examined the relationship between the two steroid receptors and the effect of age on this relationship. We used a dual immunofluorescence labeling approach to detect those receptors. Figure 11 summarizes the data that showed a progressive decrease in the relationship between ER-α and PR starting at 12 weeks of age.

DISCUSSION AND CONCLUSIONS

In rodents, mammary gland development occurs in the postnatal female in two discrete physiological states, namely, puberty and pregnancy, adulthood being considered as a quiescent stage. In this study, we focused on the pubertal development, normal adulthood and aging of the mouse mammary gland.

At birth, the mammary gland looks like a rudimentary system of ducts. During puberty, the ductal epithelium differentiates into two layers: an inner functional parenchyma of luminal epithelial cells responsible for the passage of milk, and an outer layer of contractile myoepithelial cells functioning in milk ejection. Each duct terminates in a club-shaped terminal end bud (TEB) comprising multiple

layers of luminal cells surrounded by a monolayer of cap cells. TEBs extend into the substance of the fat pad (figure 4), resulting in ductal morphogenesis. Ultimately, the whole mammary fat pad accommodates a complex network of ducts. At 6 weeks of age the fat pad is still partially filled with numerous big TEBs and cells in division (proliferation). By 12 weeks of age, the fat pad is totally filled but some TEBs of small size can still be seen (figure 6). At 18 weeks of age the gland can be considered morphologically adult and quiescent (figure 5) with 100% filled fat pad, mature alveoli structures, no TEBs, and few or no proliferating cells. With aging, starting at 36 weeks, the gland undergoes a progressive regression of the ductal tree (figure 5) with a reduction in lateral branchings and alveoli structures, and an increase in cell death (apoptosis and macrophage invasion) within the epithelium (data not shown).

Estrogen and progesterone play a key role in all this complex process of normal development of the mammary gland (Mixner J. et al 1942).

The plasma level of estrogen increases at puberty to reach a plateau at adulthood (Danilovich N. et al. 2003). It is well established that ER-α is implicated in ductal growth (Boccinfuso W.P. et al. 1997) and PR is important for lateral branching (Shyamala G. et al. 1999). This would explain why we found that the number of positive cells for both markers reaches its maximum at puberty (figure 10) and progressively diminish to reach a plateau at 18 weeks of age when the gland is morphologically developed. At this stage the gland is considered quiescent.

The expression of these receptors is heterogeneous within and between different morphological structures, and in different stages of the development of the animals. We believe that the distribution, and not only the number of cells expressing the receptors has an impact in the normal –or abnormal- development of the gland, since they are a manifestation of the delicate homeostatic equilibrium of the gland. This level of heterogeneity can be completely missed when looking at total protein contents of the gland. In addition, accurately measure the whole-gland distribution can not be done using standard 2D microscopy because the complexity of the gland and the effects of sectioning

might lead to inaccurate classification of tissue structures or cells. Due to this factor, small sections of large ducts or end buds can be easily classified as small ducts and/or myoepithelial cells can be accounted for as luminal epithelial cells. Thus we have used our in-house developed 3D reconstruction system to reconstruct the glands from physical sections and quantify the receptor expression in the entire 3D extent of the gland. The choice of physical over optical –confocal- sectioning, which would provide better Z resolution and registration between sections, obeys to the thickness of the tissue, which would impede both staining and optical sectioning trough the entire thickness of the tissue.

To better understand the relationship between mammary gland anatomy and receptor expression, nuclei in large ducts, small ducts and TEBs were quantified separately (figure 10A and 10B). At puberty ER-α is significantly higher in small ducts and end buds where ductal morphogenesis occurs. PR, however, is important for lateral branching and is higher in small ducts. During puberty, both receptors decrease progressively and changes are more dramatic in small ducts than large ducts. Those results confirm the role of those receptors in the normal development of the gland (Boccinfuso W.P. et al. 1997, Shyamala G. et al. 2002).

The mice never outlive their reproductive capacity but estrogen plasma levels decrease with aging (Danilovich N. et al. 2003) and since estrogen downregulates ER-α (Shyamala G. et al. 2002), we would expect the level of expression of ER-α to increase. Surprisingly, although no dramatic changes in the number of positive cells were observed, the level of expression of ER-α, determined by the intensity of staining, decreased noticeably. This might be explained by the fact that estrogen sensitivity of the mammary gland (Shyamala G. et al 2002) increases with aging.

Previous studies have shown that PR is usually coexpressed with ER-α in non-proliferating mammary epithelial cells (Clarke R.B., 1999) and that PR expression is regulated directly by progesterone and indirectly by estrogen through ER-α (Shyamala G. et al. 1997). Here, we showed that although ER-α and PR are

highly colocalized, the percentage of colocalization decreases with the age of the animal (figure 11) and the level of PR expression and the number of PR positive cells are constant through adulthood (figure9). A possible biological explanation of these results would be that after puberty the remaining PR in the adult gland is more likely to be constitutive, independent of ER-α.

In Summary, we have presented a methodology and tools for the 3D histopathological study of the mammary gland. We have shown how, using a computer assisted microscopy system, the entire mammary gland of a mouse can be reconstructed from consecutive tissue sections, and how the presence and distribution of specific molecules can be quantified using this reconstruction. We believe that this work is a technical accomplishment that will have an important impact in understanding the role of hormones in normal development and carcinogenesis. We were able to classify the structures properly, to follow them through though the sections and to reconstruct them. We have analyzed very high number of nuclei (1200-12000 nuclei/class/case). We could also store the data in a convenient way that relates the expression analysis to the morphology of the tissue. We have compiled and combined an enormous amount of information regarding both morphology and distribution of molecular markers. We believe that this approach provides more accurate and objective results than those obtained with other more manual methods which, as previously described, are sometimes hampered by methodological issues (proper tissue and cell classification) and human biases.

ACKNOWLEDGEMENTS

The research described here was performed at the Lawrence Berkeley National Laboratory and funded by the Department of Energy Office of Science under Contract No. DE-AC03-76SF00098, the California Breast Cancer Research Program under project number 8WB-0150, and the Department of Defense Breast Cancer Research Programs under project number DAMD17-00-1-0306.

References

Boccinfuso W.P., Korach K.S., 1997, Mammary gland development and tumorigenesis in estrogen receptor knock-out Mice, Journal of Mammary Gland Biology and Neoplasia, 2(4): 323-344.

Cheng G., Weihua Z., Warner M., Gustafsson J.A., 2004, Estrogen receptors ER alpha and ER beta in proliferation in the rodent mammary gland, Proc Natl Acad Sci U S A. 101(11): 3739-46.

Clarke R.B., Steroid receptors and proliferation in the human breast, Steroids, 68(10-13): 789-794.

Danilovich N., Harada N., Sairam M.R., Maysinger D., 2003, Age-related neurodegenerative changes in the central nervous system of estrogen-deficient follitropin receptor knockout mice. Exp Neurol., 183(2): 559-72.

Dean, P., Mascio, L., Ow, D., Sudar, D., Millikin, J., 1990, Proposed Standard for Image Cytometry Data Files. Cytometry 11: 561-569.

Fernández-Gonzalez R., Jones A., Garcia-Rodriguez E., Chen P.Y., Idica A., Barcellos-Hoff M.H., Ortiz-de-Solórzano C. 2002, A system for combined three-dimensional morphological and molecular analysis of thick tissue samples. Microscopy Research and Technique 59(6): 522-530.

Lazannec G., Bresson D., Lucas A., Chauveau C., and Vignon F, 2001, ERβ inhibits proliferation and invasion of breast cancer cells, Endocrinology, 9(142): 4120-4130.

Levin E.R., 2001, Cell localization, physiology, and non-genomic actions of estrogen receptors, Journal of Applied Physiology, 91(4): 1860-1867.

Lydon J.P., DeMayo F.J., Funk C.R., Mani S.K, Hughes A.R., Montgomery Jr. C.A., Shyamala G., Coneely O.M., O'Malley B.W., 1995, Mice lacking progesterone receptor exhibit pleiotropic reproductive abnormalities, Genes and Development, 9(18): 2266-2278.

Mixner J. and Turner C., 1942, Role of estrogen in the stimulation of lobule-alveolar growth by progesterone and by the mammogenic lobule-alveolar growth factor of the anterior pituitary, Endocrinology, 30: 591-597.

Ohtake T., Abe R., Kimijima I., Kukushima T., Tsuchiya A., Hoshi K., Wakasa H., 1995, Intraductal extension of primary invasive breast carcinoma treated by breast-conservative surgery. Computer graphic three-dimensional reconstruction of the mammary duct-lobular systems. Cancer 76(1): 32-45.

Ohtake T., Kimijima I., Fukushima T., Yasuda M., Sekikawa K., Takenoshita S., Abe R., 2001, Computer-assisted complete three-dimensional reconstruction of the mammary ductal/lobular systems: implications of ductal anastomoses for breast-conserving surgery, Cancer, 91(12): 2263-72.

Ohuchi N., Furuta A., Mori S., 1994, Management of Ductal Carcinoma In Situ with Nipple Discharge. Cancer 74: 1294-1302.

Shyamala G., Barcellos-Hoff M.H., Toft D., Yang X., 1997, *In situ* localization of progesterone receptors in normal mouse mammary glands: absence of receptors in the connective and adipose stroma and a heterogenous expression throughout the epithelium, Journal of Steroid Biochemistry and Molecular Biology, 63(4-6): 251-259.

Shyamala G., 1999, Progesterone signaling and mammary gland morphogenesis, Journal of Mammary Gland Biology and Neoplasia, 4(1): 182-190.

Shyamala G., Chou Y.C., Louie S.G., Guzman R.C., Smith G.H., Nandi S., 2002, Cellular expression of estrogen and progesterone receptors in mammary glands: regulation by hormones, development and aging, Journal of Steroid Biochemistry and Molecular Biology 80: 137-148.

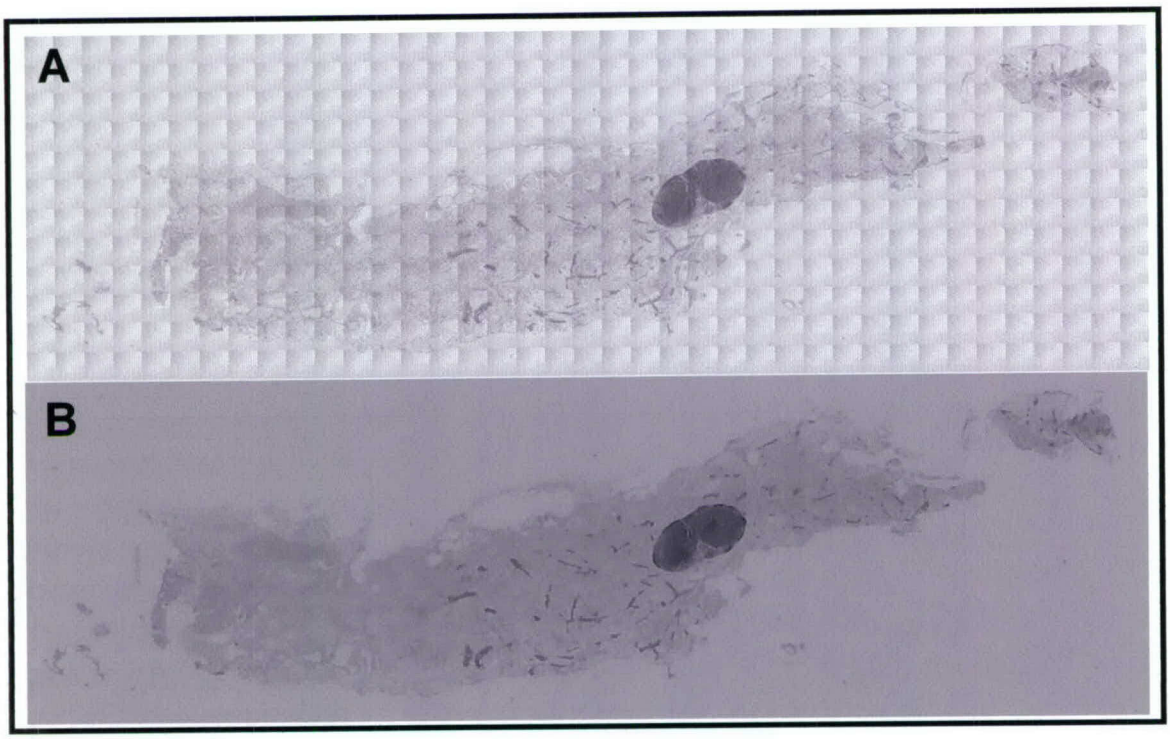


Figure 1. Low resolution scans of an H&E-stained section A) before and B) after background correction.

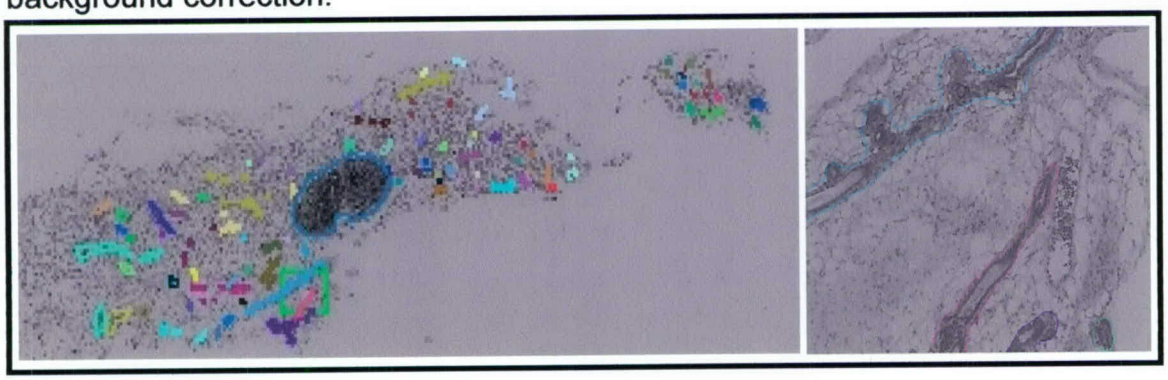


Figure 2. Low resolution scans of an H&E-stained section after background correction and segmentation. The colored shapes delineate the epithelial structures and the lymph node.

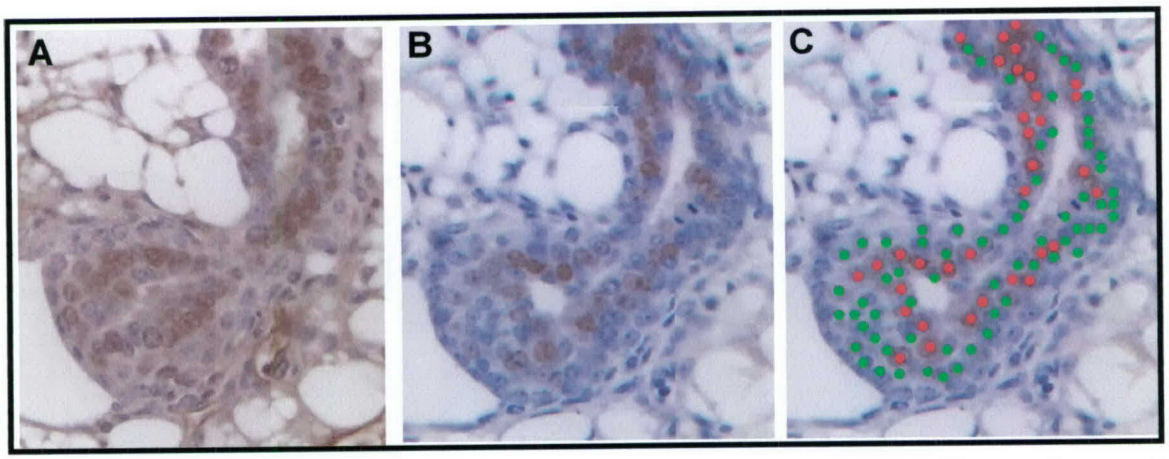


Figure 3. ER staining (A) on a small TEB and PR staining (B) on the next consecutive section. Dark brown cells represent positive nuclei and blue-purple cells represent negative nuclei. C/ Interactive markings of PR staining. Red dots represent positive nuclei and green dots represent negative nuclei.

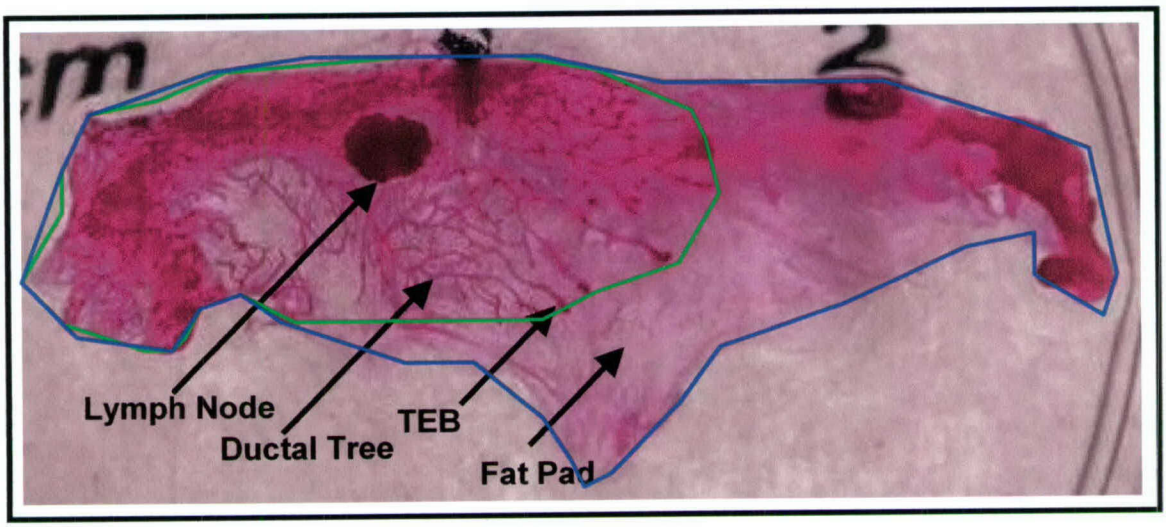


Figure 4. Whole mount of 6-week inguinal mammary gland. Ductal growth is incomplete, and several end buds can be seen extending toward the periphery of the fat pad. The green line surrounds the epithelial structures and the blue line surrounds the fat pad. The percentage of filled fat pad is defined by the surface of the epithelium by the surface of fat pad.

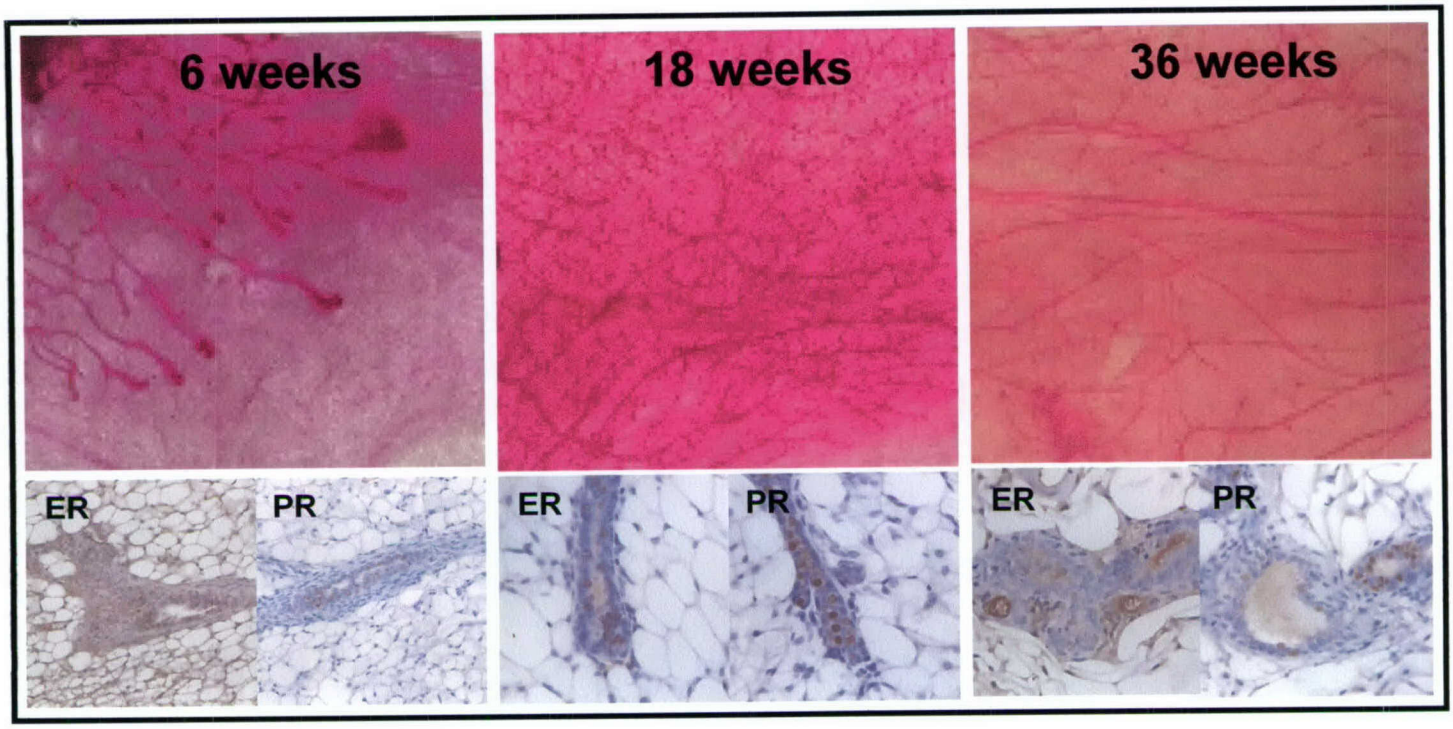


Figure 5. Whole mount slide show (up) of portions of inguinal mammary glands at 6, 18 and 36 weeks showing how the gland develops (6 and 18 weeks) and regresses (18 and 36 weeks). Analysis of for ER and PR expression (down) at different stages.

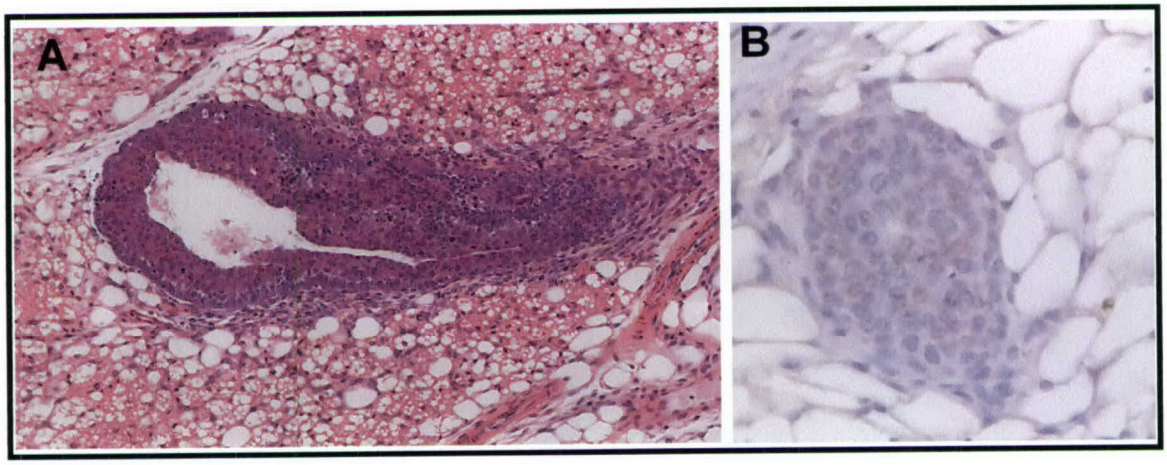


Figure 6. Images of a TEB at 6 (A) and 12 (B) weeks of age.

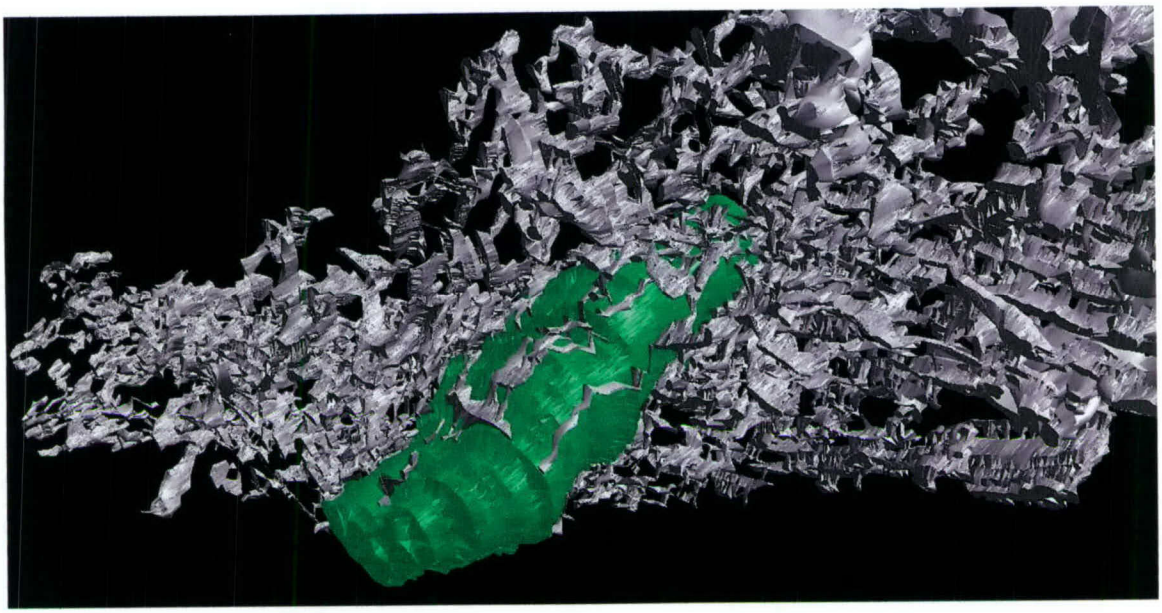


Figure 7. 3D reconstruction of the 6-week-old mouse. Normal ducts are rendered as gray volumes and the lymph node as green volume. The entire scene can be rotated and zoomed in and out. All volumes can be selected individually. The scene was stretched 10X in the Z direction to obtain a better view.

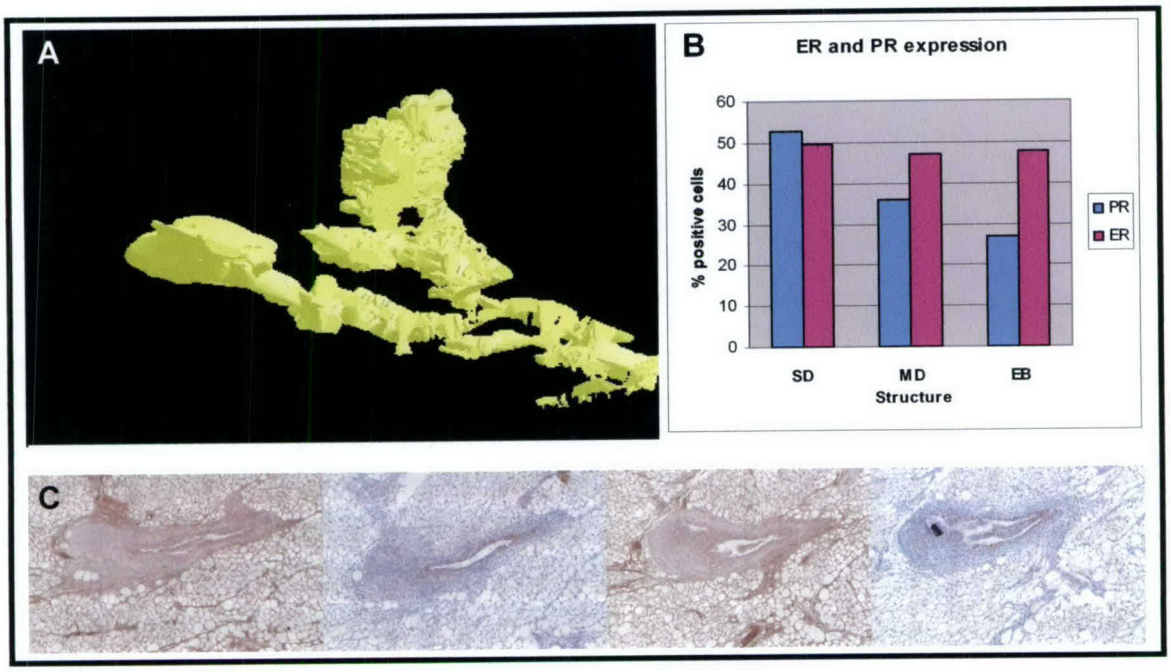


Figure 8. A/ 3D reconstruction of one duct branching into two TEBs of the 6-week-old mouse. B/ ER and PR expression in this selected structure. C/ images of the same TEB in consecutive sections.

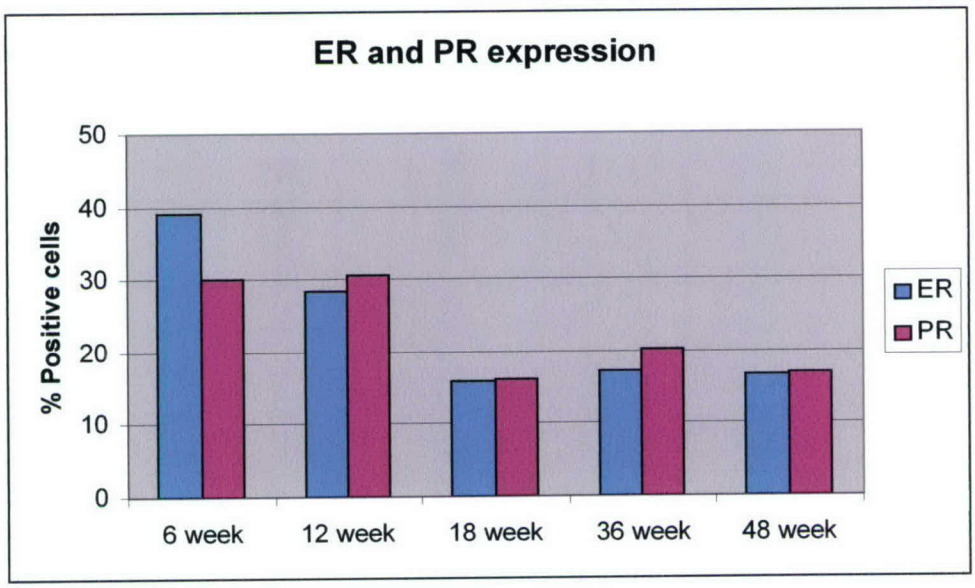


Figure 9. Modulation of ER-α (blue) and PR (red) positive epithelial cells in mammary glands during normal development (6, 12 and 18 weeks) and adulthood (18, 36 and 48 weeks).

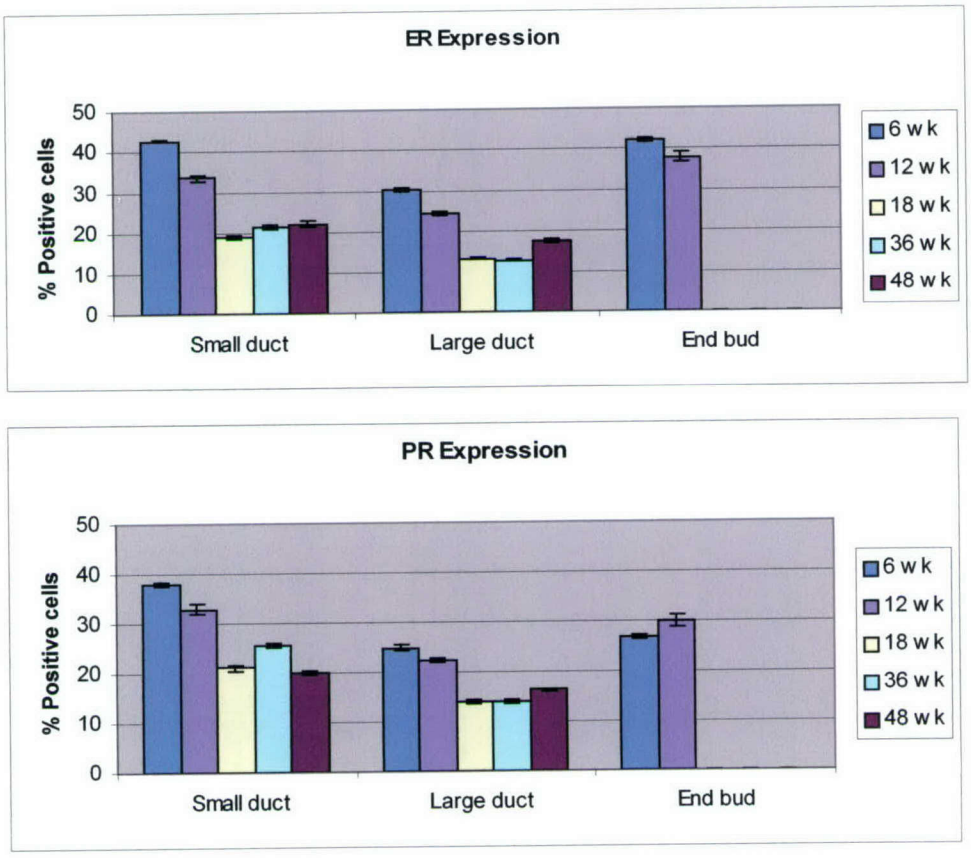


Figure 10. Modulation of ER-α (up) and PR (down) positive epithelial cells in different structure of the mammary glands during normal development (6, 12 and 18 weeks) and adulthood (18, 36 and 48 weeks).

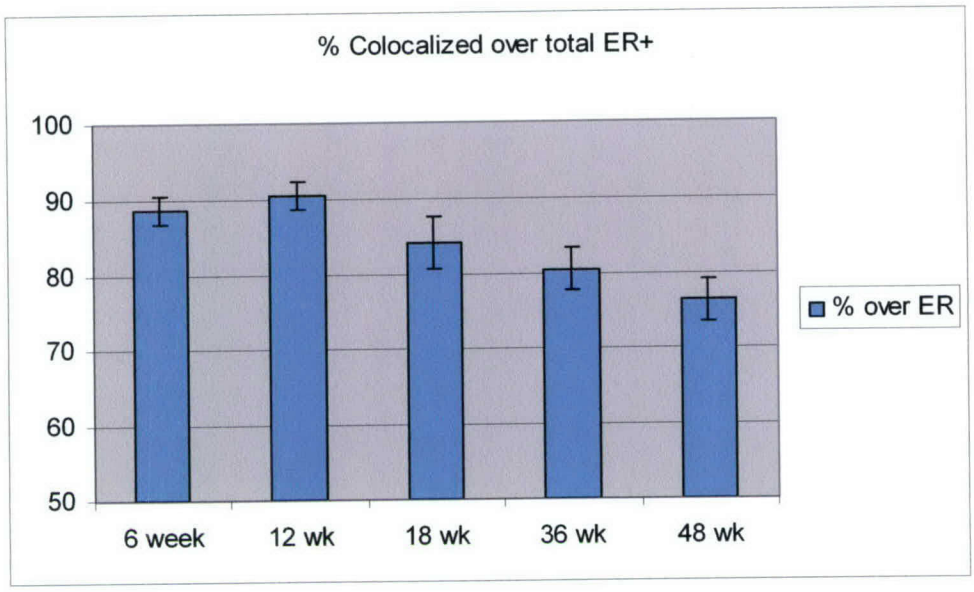


Figure 11. Quantitative analysis of ER-α and PR colocalization in nulliparous mice of different ages expressed as a percent of colocalized markers over ER-α positive cells.